Marketing Data Mining Classifiers: Criteria Selection Issues in Customer Segmentation

Masoud Abessi
Department of Industrial Engineering,
Yazd University, Iran

Elahe Hajigol Yazdi
Department of Industrial Engineering,
Yazd University, Iran

ABSTRACT
Data mining is automated or semi-automated Knowledge Discovery from large amounts of stored data in order to discovering meaningful patterns and rules. Marketing related data mining applied to market segmentation, customer services, credit and behavior scoring, and benchmarking. There are different classifiers including decision tree, ID3, CART, Quest, Neural networks, Association, Bayesian, and etc. In this study, ten classifiers are examined to identify important issues in mining marketing data. Classification accuracy, learning speed, Classification speed, Missing value, and robustness are some indices included to compare and contrast the classifiers. Shopping malls' consumer behavior data were used in our investigation. Results indicate that classifiers perform differently under different consumer data types.

Keywords
Data mining, Supervised learning algorithm, classification techniques

1. INTRODUCTION
Knowledge discovery from marketing data is an increasingly important phenomenon. Classification techniques can learn from consumer behavior, market segmentation, and customer purchasing habit to promote Customer loyalty and profitability, and products/services Customization (Gloy, Akridge, & Preckel, 1997). Marketing classifiers in a supervised learning algorithm format helps to discover distinctive groups based on certain preferences. In this study, we examine different classifiers to identify various market segmentation based on observed customer purchasing and product usage behavior, socioeconomic factors, demographic, and life stage characteristics. It creates specific opportunities to plan a strategic roadmap which` leads to the greater customer loyalty, shorter customer purchase cycles, higher spend, and lowering service and support costs (Garofalakis et al, 2000).

Various segmentation techniques exist to classify costumers according to their unique characteristics, each with their own unique advantages and challenges. Segmentation is a multi-disciplinary method which combines statistics and machine learning techniques. Statistic approach is implemented for collecting, classifying, summarizing, organizing, analyzing, and interpreting data. Statistical approaches like regression classifier can't extract tacit knowledge which is hidden in datasets. Legitimately, they cannot generate rules. Therefore, this approach doesn't have learning ability. Various studies were implemented statistical segmentation methods. Gilboa (2009) presented a two-step cluster analysis based on customer's socio-demographic characteristics to segment Israeli mall customers as disloyal customers, family bonders, minimalists, and mall enthusiasts. Wu (2006) implemented factor analysis and Discriminant classifier based on the lifestyle and personality characteristic Factors to predict the intentions and behavior of different Internet bookstores customer groups. Regression analysis applied to figure out the relationship between service quality perceptions and satisfaction and intention by lee et al (2011). On the other hand, data mining techniques exemplary machine learning methods act based on finding structure in the dataset. Kim et al. (2006) utilized decision tree with the aim of customer's identification. Whereas, target customer analysis take placed by using decision tree (Chen et.al, 2003; Wu et.al, 2005). Liang (2010) analyzed customer value by data mining technologies Integration as SOM and decision tree for the automotive maintenance industry.

According to the literature review, vast approaches were implemented for costumers' segmentation without any harmony among them. Most studies sufficed to statistical analysis of customer's data which learning process didn't take place by it. They explore data by means of describing it based on its important aspects. Hidden knowledge in data set didn't extracted; therefore, predictive power of statistical segmentation methods is low. Studies which used DM techniques for customer grouping often have a special attention to results achieved from segmentation and its characteristics. They didn't attend to model deployment according to learning process, predictive power, and knowledge tacit in extracted rules. The results of the KDnuggets 2009 Data Mining Deployment Poll indicate that more than 55% of people who used machine learning techniques didn't deploy the model and just utilized the results to gain business, scientific knowledge, or publish papers. Various Learning methods work under their unique conditions. Another problem with segmentation model selection is that the customers data are normally stored in various forms including binary values (asymmetric or symmetric), ordinal or ratio scales depending on marketing subjects. For example many customers' data are collected in Likert scale or other ordinal forms from questionnaires filled by consumers (Michael & Gordon, 1997). Whereas, different validation techniques are implemented various evaluation criteria. They are including Predictive (Classification) accuracy, speed, robustness, Scalability, Interpretability, Simplicity and Domain-dependent quality indicators (Yeh &Lien, 2009; Ture & et.al, 2009). Therefore, the admissible classification method varies according to the data characteristics and business requirements (Carrier & Povel, 2003).

This paper aims to represent a framework to facilitate classifier selection for customer segmentation according to data characteristics and business requirements, legitimately. First, classification models and their characteristics presented. We refer to their strengths and limitations (Kims 2008). Critical factors in classification method selection which is applicable to marketing data proposed for ambiguous reduction in technique selection. We apply various classification models to a shopping malls data set to classify shopping malls customers according to their demographic and psychographic characteristics, consumer purchasing behavior and customer expected benefit from her/ his shopping. Finally, a systematically integrated roadmap develop for classification technique selection based on data type and business requirements by considering obtaining results of evaluation criteria such as robustness, accuracy, speed and etc.

2. CLASSIFICATION MODELS

Different approaches exist for Classifiers grouping. For example, classification methods classified to logical, cross-tabular, and equation-base techniques. Furthermore, in another approach classifiers were assorted as decision tree algorithms, stochastic, eager, lazy, and heuristic classifiers. In this study, we implemented the most common classification methods for customer segmentation as decision tree algorithms, K-nearest neighbor, Logistic regression, Naïve Bayesian, association, SVM, and artificial neural networks classifiers.

Decision tree algorithms (Quinlan, 1993) rooted from a nonparametric approach which doesn't require any prior assumption about classes and data distributions or scales (Berry &Linoff, 2000). Decision tree algorithms lead to a set of "if-then" rules to improve human readability. It uses discrete value function that is robust to the noisy data. It is capable to handle "or" expression learning. DT implements ordinal and nominal data as input variables. One of the major obstacles with continuous data occurs when over-fitting take places. Decision tree approach is robust to errors in classification of the training set. The presence of redundant attributes does not have a nugatory affect on the decision trees accuracy. Some decision tree methods can handle missing values such as C4.5 while ID3 can't. Decision tree algorithms never backtrack to reconsider earlier choices. DT method splits the data by single variable at a time which can cause a sub-tree replication and high error rates. Optimal decision tree attainment is a NP-hard problem. Decision tree construction is jitney. The resulted tree is expressive interpretable especially for small size trees. There are different DT methods such as ID3 (Quinlan Ross,1986), CART (Breiman, 1984), CHAID (Kass, 1980) and C4.5 (Quinlan, 1993) and QUEST (Loh and Shin,1997) which are implemented for customer segmentation. They are differing according to their knowledge extraction methods as well as their accuracy, speed of classification, missing values, noisy data handling, and etc. In this study, we use general models of ID3, CART, CHAID, C4.5 and QUEST with 5 tree depth for customer segmentation.

A Bayes classifier (M. Maron, 1961) apply Bayes' theorem which implement Maximum likelihood method for parameter estimation (Bhargavi and Jyothi, 2009). Bayes classifier requires small amounts of training data for classification parameters estimation. It only uses categorical data for classification. The predictive accuracy is highly correlated with the assumption of class conditional independence which simplifies computation (Delen et al, 2004). Na¨ıve Bayes Classifier is a proper method of Bayes classifiers which is robust to isolated noisy points as well as irrelevant attributes. It can handle missing values well by means of the example ignorance in model building process. It is robust to over fitting Because of existence of prior knowledge.

Associative classification model as a rule-based classifier groups data by discovering associations between a set of features and a class label (jalali, 2010). Associative classifier tuning is so hard. A large minimum support may lead to choose only rules which are contained obvious knowledge and don't consider exceptional interesting tacit knowledge. A low minimum support consideration causes huge numbers of redundant or noisy rules. In this research, we consider minimum antecedent support up to 10% and minimum rule confidence sets to 90%. The association classifier is equivalent to decision tree according to expressiveness which builds rectilinear partition from attribute space. It construct descriptive model which is easier to interpret.

Logistic regression (Hastie, 2001) as a special case of linear regression models accepts both continuous and categorical data as input format types. The major advantage of this approach is that it can produce a simple probabilistic formula of classification whereas logistic model is well understood. The weakness of logistic regression is linearity assumption of response variable according to the coefficients of the predictor variables. Logistic regression cannot properly deals with the problems of non-linear and interactive effects of explanatory variables (Delen et al, 2004).

The k-nearest neighbor method (T.Cover & P. Hart ,1950s) is an instance-based learner. KNN method usage for large training data sets do not require model Building which lead to huge calculation. Speed of classification in KNN technique as a lazy classifier is low. It accepts only continuous and categorical data and work under static scenario affected by sample size. The major advantage of this approach is that it is not required to establish predictive model before classification. KNN does not produce a simple classification probability formula and its predictive accuracy is highly affected by the measure of distance and the cardinality k of the neighborhood (Yeh and Lien, 2009).

Support vector machines (Vapnik and Cortes, 1995) are set of related supervised learning methods which rooted from statistical learning theory, implemented for classification and regression analysis. The SVM takes a set of input data, and predicts based on hyper-plan maximum margin. The SVM advantage is the model only deals with these support vectors, rather than the whole training dataset. Then, the size of the training set is not usually an issue. Support Vector Machines have been found to perform well on problems that are non-linear, sparse, and high dimensional. A special property of SVM classifier is that they simultaneously minimize the empirical classification error and maximize the geometric margin. SVM classifier is sensitive to the choice of variable settings, making it harder to use, and time consuming to identify the best. SVM learning problem is a convex optimization one which facilitates global optimum finding while other classification methods such as rule based classifiers and NN employ greedy based strategy.

Artificial neural networks (McCulloch and Pitts, 1943) have developed from generalized neural biological principles. Neural network involves long training times. Therefore, it is more suitable for feasible applications. The accuracy of resulted classes is high. Neural networks have been criticized for their poor interpretability. It is difficult for human to interpret the symbolic meaning behind the learned weights and "hidden units" in the network. The advantages of neural networks include their high tolerance of noisy data as well as their ability to classify patterns on which they have not been trained. Redundant attributes can handle by NN because automatic learning of weights. To avoid model over-fitting, a proper network topology should be chosen. They can be used when little knowledge of the relationships between attributes and classes exist. They are well-suited for continuous-valued inputs and outputs, unlike most decision tree algorithms.

Neural networks work under both static and dynamic scenarios.

In next section, the proposed classifiers applied for shopping malls customers' segmentation and compare their performances.

3. EXPERIMENTAL RESULTS

In this research, we aim to extract knowledge from customers' data by means of various classifiers to group customers who referred to different shopping malls. Customer segmentation was took place according to their demographic characteristics, psychographic elements which have impact on customer purchasing behavior and benefits that customers expect from their shopping. The data gathered from 698 respondents which referred to 11 shopping malls, to know who buy what in which shopping mall. Feature selection module was applied which lead to the most proper variables selection with the most critical impacts on the classification problem.

The research variables designed to measure Quality, brand preferences, customer loyalty, mall environment, consumer psychographic characteristics, customer Motivation for mall visits, favorite activities, their sensation about purchasing and shopping mall purchasing behavior. These variables Sketched in 4 groups of factors consisting of psychographic factors, consumers behavior factors, benefit factors, and socio-demographic variables. Internal consistency (Cronbach, 1951) of achieved factors measured by cronbach's alpha is 0.78. So, from 108 variables the 32 most important variables choose for applying by classification techniques.

Table 1- socio-demographic characteristics of costumers

socio-demographic	N	Percent (%)	socio-demographic	N	Percent (%)
Gender			Occupation status		
Men	302	43.27	Full time	371	53.15
Women	394	56.45	Part time	54	7.74
Age			student	249	35.67
Range of age	1939-	--	House wife	13	1.86
Average age	1991	--	Ethnic		
Marital status			Malay	358	51.29
Single	521	76.64	Chinese	220	31.52
Married without child	48	6.88	Indian	99	14.18
Married with	121	17.34	others	20	2.87
Divorced and widowed	8	1.15	income		
Education			Below average	472	69.51
Under diploma	231	31.1			
diploma	161	23.07	Average	99	14.58
B.S degree	257	36.72			
Post graduate	42	6.02	Above average	108	15.90
others	7	1			

The shopping malls customers' database (as table 1) was composed of Forty four percent (302) of the male respondents while fifty six percent (394) of the respondents were female customers. Among them 358 was Malaya (51.29%), 220 (31.25%) was Chinese, and 99(14.18%) was Indian. The interviewees' birth year range from 1939 to 1994. The most customers' birth year range between 1983 and 1991 with the most frequency in 1991. From interviewee 521 were single (74.64%), 48(6.88%) were married without child, 121(17.34%) were married with children, and 8 (1.15%) were divorce or widow. Regarding education status, 231(31.1%) had under diploma degree, 161 (23.07%) were diploma, 257 (36.82%) have B.S , and 42(6.02) had postgraduate

degree.371(53.15%) were fulltime workers,54 (7.74%) were part time workers, 249 (35.67%) were adolescents or students ,and 13 (1.86)were housewives. The respondents level of income reported 472 customers (69.51%) were below average, 99 (14.58%) were average, and 108 (15.90%) were above average. Customers' group identification took place by means of various classification techniques. In this study, we applied C5.0, CART, QUEST, CHAID, ID3, Logistic Regression, Bayesian network, SVM, Neural Network, and Association classifier. We run these models in SPSS Clementine 11.0 software. Data set was built from both categorical and continuous data. Evaluation results based on classifiers accuracy and their maximum profit rates are presented in table 2. Applied models rank according to their accuracy in classification.

Table 2- accuracy- Maximum profit of classification models

Method	C5.0	CART	SVM	NN	CHAID	QUEST	BEYS	ID3	LR	DISCR
Accuracy	92.86	92.55	92.28	92.19	91.70	91.44	90.64	90.55	90.40	82.52
Max profit	4906	4720	4535	4470	4145	3967	3550	3290	3305	3420

As presented C5.0, CART, SVM and NN are more accurate. According to maximum profit gained by different models and provided accuracy from decision tree algorithm, C5.0 and CART methods are more efficient. As mentioned various classifiers have different performances when applied on shopping malls customers' data. The manner of Knowledge extraction, the level of variable usage and the rules they implemented for prediction are critical elements which made them have different performance on the same database. For example, CART algorithm ran to segment customers. The knowledge hidden in data base extracted by means of 15 rules while CHIAD method implemented 11 rules to customers segments prediction. Whereas, the number of variables used for building tree are different in CART and CHAID algorithms. Various classifiers need different level of data requirement. Some methods use all of data while other doesn't. C5.0 classifier used all variables specified by means of feature selection which extract ten best variables are most important to accurate classification. Discriminant analysis implemented all of variables to make a linear function but namely, a priori classifier only used socio-demographic variables to extract 765 rules. Therefore, Various Learning methods work under their unique conditions. Statistical approaches like LR and Discriminant analysis don't extract tacit knowledge from database. They don't have learning ability. While, DT algorithms utilize learning process, predictive power, and knowledge tacit in extracted rules. The richness of the rules gives this technique the potential of reflecting the true classification structure in the data. Table 3 shows knowledge extraction construction of various classifiers.

Table 3- knowledge extraction construction of rule-based classifiers

	C5.0	CART	CHAID	QUEST	BEYS	DISCR	LR
N. variable	32	14	7	5	7	32	32
Tree level	8	9	4	1	--	--	--
N. rules	20	15	11	2	765	--	

Decision tree algorithms works under Static scenarios have good ability in knowledge representation, since the tree derived from the model have a very straightforward interpretation unlike the obscure models as NN. Results from Bayesian classifier is in IF-Then rules which facilitates knowledge representation latent in the database. But, the ecstasy number of extracted rules against the little number of implemented variables made it inefficient on this problem. Learning speed of Decision tree methods is high. So, these methods can classify database quickly. As presented in table 3 SVM prepare a noticeable profit among other methods. SVM is particularly suited for using with wide datasets. Most of classifiers as SVM are sensitive to starting variables. Neural Network requires minimal statistical or mathematical knowledge to train or apply whilst classify database with high accuracy and acceptable profit. Association classifier could not generate efficienal model from current database because of high threshold presence. So, we obliged to implement a huge dataset. Association classifier capability in handling miss value, noisy data and duplicate cases is acceptable. Despite of learning speed, the speed of classification is high. It's important to know Support vector machine (SVM) technique and discriminate method have considerable less learning speed while Bayesian, KNN and decision tree are too fast in learning process. On the other hand, all classification methods except lazy methods (such as KNN and CBR) are fast in classification. According to resulted from running different models on proposed database table 5 shows the results of applying different techniques according to various model parameters.

Table 4- classification models comparison

Method / Evaluation criteria	Bayesian	Decision tree	KNN	Association classifier	Neural Network	SVM
Accuracy	low	medium	medium	medium	high	Very high
Learning speed	Very high	high	Very high	medium	low	low
Classification speed	Very high	Very high	low	Very high	Very high	Very high
Miss value robust.	Very high	high	low	medium	low	medium
Noisy value robust.	medium	high	medium	medium	low	Very high
Duplicate robustness	low	medium	medium	medium	medium	high
Dependent char. robustness	low	medium	low	medium	high	high
Data type robustness	high	Very high	high	high	high	high
Over fitting	high	medium	high	medium	low	medium
Incremental learning	Very high	medium	Very high	low	high	medium
Knowledge represent	Very high	Very high	medium	Very high	low	low
Param. management	high	high	high	high	low	low

Kinds of data stored in database plays a critical role in appropriate model selection for marketing data. Some model executes with categorical data while other are incurious to data type. For example, NN are apathetic to data type whereas lazy classifier such as KNN and CBR are sensitive to data type which implement multi model distribution or sample size data. SVM and NN methods are more efficient in applying multi dimension-continues variables. On the other hand Logic Based Systems acts better in discrete categorical attributes. Bayesian classifier, Discriminant classifier and Log regression model can handle any type of data as input whereas Target variables must be categorical. Statistical models such as log regression have capability in handling non normal and non linear data with high correlation.

Discriminant analysis and Logistic Regression are both suitable classification models. Log regression model is often quite accurate. But Discriminant analysis produces better results especially in the case of small sample size. Association rules can handle numeric input data and multiple target variables. They generate more general rules than resulted from decision trees because of rules overlap property in the rule set. Through Decision tree algorithm, C5.0 and QUEST can handle any type of data as input variable but the target variable must be categorical. Whereas, CHAID and CART algorithms can accommodate continues or categorical data. Its noticeable ordinal fields used in these models must have numeric storage. CART and C5.0 models are quite robust when faced with the problem of missing data, noisy data and large numbers of input fields. Unlike CART and QUEST models, CHAID can generate non binary trees. CHAID and QUEST implement statistical test to estimate what predictor must be used. Impurity-change measure applied by QUEST and CART to predictor variable selection and split determination. Also, Accuracy promotion happens by applying a powerful boosting method in C5.0. Table 4 acts as a strategic roadmap which facilitates classifier selection when we have enough knowledge about marketing data types.

Table 5- data characteristics/ various methods (CA= Categorical data, C=Continues data, B=both data kinds)

Method / Evaluation criteria	CHAID	CART	C5.0	QUEST	BAYS	KNN	DISC	LR	AC	NN	SVM
predictor variable	B	B	B	B	CA	B	B	B	B	B	B
Target variable	B	B	CA	CA	CA	B	CA	CA	B	B	CA

4. CONCLUSION

The proper classification technique selection for knowledge extraction from marketing data is one of the most critical tasks. We study different classification techniques that are applicable to marketing data and their strength and weakness which present in Appendix A briefly. It prepares a comprehensive view on different classification techniques. Specified hypothesis identification of each classification methods help us to select the most profitable classification technique which can be applied to marketing problem. Also, data type recognition is crucial to identify specified classification methods that are suitable in choosing a proper classification technique. Future research can focus on promotion different classification methods with the aim of models restriction dominance such as study various solution to overcome classification models shortcoming in handle different data types

5. REFERENCE

[1] A. Al-adaideh. Q, the Impact of Classification Evaluation Methods on Rough Sets Based Classifiers.

[2] Bhargavi, P and Jyothi, S. (2009). Applying Naive Bayes Data Mining Technique for Classification of Agricultural Land Soils. International Journal of Computer Science and Network Security, VOL(9) No 8.

[3] Breiman, L., Friedman, J., Olshen, L and Stone, J. (1984). Classification and Regression trees. Wadsworth Statistics/Probability series. CRC press Boca Raton, Florida, USA.

[4] Chen, J., Xing, Y., Xi, G., Chen, J., Yi, J., Zhao, D., Wang, J. (2007) A Comparison of Four Data Mining Models: Bayes, Neural Network, SVM and Decision Trees in Identifying Syndromes in Coronary Heart Disease, In: D. Liu et al. (Eds.): ISNN 2007, Part I, LNCS 4491, pp. 1274–1279.

[5] Garofalakis, M., Hyun, D., Rastogi, R. and Shim, K. (2000). Efficient algorithms for constructing decision. trees with constraints. Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 335 – 339.

[6] Gilboa S., (2009) ,A segmentation study of Israeli mall customers, Journal of Retailing and Consumer Services 16, 135–144.

[7] Han, J., and Kamber, M. (2006). Data Mining: Concepts and Techniques, Second Edition. Elsevier.

[8] Hunt, E.B., Marin. and Stone,P.J. (1966). Experiments in induction, Academic Press, New York.

[9] Kohavi. R, (1995), A Study of Cross Validation and Bootstrap for Accuracy Estimation and Model Selection, International Joint Conference on Artificial Intelligence.

[10] Kotsiantis, S.B. (2007) “Supervised Machine Learning: A Review of Classification Techniques”, Informatica, Vol. 31, pp. 249-268.

[11] Khoshgoftaar, T.M and Allen, E.B. (1999). Logistic regression modeling of software quality. International Journal of Reliability, Quality and Safety Engineering, vol. 6(4, pp. 303-317.

[12] Lee J., Kim H., Ko Y., Sagas M., (2011) , The influence of service quality on satisfaction and intention: A gender segmentation strategy, Sport Management Review 14, 54–63.

[13] Lessmann S., VO S., (2009), A reference model for customer-centric data mining with support vector machines, European Journal of Operational Research 199, 520–530.

[14] Lewis, R.J. (2000). An Introduction to Classification and Regression Tree (CART) Analysis. Annual Meeting of the Society for Academic Emergency Medicine, Francisco, California.

[15] Lian Yi-Hui, (2010), Integration of data mining technologies to analyze customer value for the automotive maintenance industry, Expert Systems with Applications.

[16] Lippmann, R. (1987). An Introduction to computing with neural nets. IEEE ASSP Magazine, vol. 22.

[17] Loh, W.-Y., & Shih, Y.-S. (1997). Split selection methods for classification trees. Statistica Sinica, 7, 815–840.

[18] Matthew N. Anyanwu and S.ajjan G. Shiva. (2010). Comparative Analysis of Serial Decision Tree Classification Algorithms. International Journal of Computer Science and Security, (IJCSS) Volume (3) : Issue (3).

[19] Mehta, M., Agrawal, R., and Rissanen, J. (1996). SLIQ: A fast scalable classifier for data mining. In EDBT 96, Avignon, France.

[20] Michael, J. A., Gordon, S. L. (1997). Data mining technique: For marketing, sales and customer support. New York: Wiley.

[21] Muthitacharoen A., Gillenson M., Suwan N., (2006) , Segmenting online customers to manage business

resources:A study of the impacts of sales channel strategies on consumer preferences, Information & Management 43, 678–695.

[22] Ngai E.W.T., Xiu L., Chau D.C.K., (2009), Application of data mining techniques in customer relationship management: A literature review and classification, Expert Systems with Applications 36, 2592–2602.

[23] Othman, M.F.B., Yau, T.M.S. (2007) Comparison of Different Classification Techniques Using WEKA for Breast Cancer, In: F. Ibrahim, N.A. Abu Osman, J. Usman and N.A. Kadri (Eds.): Biomed 06, IFMBE Proceedings 15, pp. 520-523, 2007.

[24] Podgorelec, V., Kokol, P., Stiglic, B., Rozman, I. (2002). Decision trees: an overview and their use in medicine, Journal of Medical Systems Kluwer Academic/Plenum Press, vol.26, Num. 5.

[25] Quinlan, J. R. (1986). Induction of decision trees. Machine Leaning, vol (1), pp.81-106.

[26] Quinlan, J. R. (1987). Simplifying decision trees, International Journal of Machine Studies, number27, pp. 221-234. Quinlan, J. R. (1993). C45: Programs for Machine Learning. Morgan Kaufmann, San Mateo, CA.

[27] Rygielski C., Wang J., Yen D., (2002), Data mining techniques for customer relationship management, Technology in Society 24, 483–502.

[28] Ture, M., Tokatli, F., & Kurt, I. (2009a). The comparisons of prognostic indexes using data mining techniques and Cox regression analysis in the breast cancer data. Expert Systems with Applications. 36 . 8247–8254.

[29] Ture, M., Tokatli, F., & Kurt, I. (2009b). Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients. Expert Systems with Applications, 36(2P1), 2017–2026.

[30] Yeh, I.-C and Lien C.-h. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Systems with Applications. 36. 2473–2480.

[31] Wu S., (2006) , A comparison of the behavior of different customer clusters towards Internet bookstores Information & Management,43, 986–1001.