Customers Satisfaction in Shipping Companies under Artificial Intelligence and Multicriteria Decision Analysis

Loukeris Nikolaos

PhD student
Technical University of Crete
Department of Electronics and Computer Engineering
Akrotiri Campus, 73100 Chania, Crete, Greece
Email: nikosloukeris@gmail.com

and

University of Essex
CCFEA
Wivenhoe Park, CO4 3 SQ, UK

Abstract:. Strategic Planning is formed considering customers satisfaction to maximise the market share. In shipping companies, the identification of satisfaction within clients is very difficult, thus satisfaction's prediction provides valuable information. Previous research in the field used techniques of multicriteria analysis, data mining and analytical-synthetical preference models. This research paper aims to define the most effective method to predict satisfaction, between techniques of data mining, rough sets, neural networks and multcriteria decision analysis.

Keywords: Shipping, Neural Networks, Multicriteria Analysis, Data Mining, Rough Sets

I. INTRODUCTION

MARITIME companies in cabotage have an intense competition that reduces Earnings Before Interest and Taxes-EBIT, thus critical improvements in every aspect of shipping operation can make the difference. Passengers opinion, maintain a very significant part as well as their final satisfaction from services. Especially in the exigent area of cruise and passengers shipping, decisions making meets vast difficulties. An interesting idea is whether prediction of a passenger's total satisfaction can be realized a priori, saving resources and maximizing corporate earnings. Previous research in the field, was carried out by Multicriteria Satisfaction Analysis (MUSA) **[1]** an analytical-synthetical method, and a combination of Multicriteria Analysis and Data Mining to predict total customers' satisfaction **[1]**. **[18]** evaluated corporate financial performance and bankruptcy prediction with Artificial Intelligence methods. The present research evaluates methods of i) Artificial Intelligence, such as Data Mining, Rough Sets, Neural Networks and ii) of Multicriteria Decisions Analysis to predict satisfaction of passengers. These methods were based on questionnaires regarding customer satisfaction at a big Greek shipping enterprise, and are quite promising in delivering acceptable results. Next paragraph describes the current situation. Paragraph 3 provides the theoretical background; in paragraph 4 are the results and conclusions are in paragraph 5.

II. CURRENT SITUATION

Decision makers are interested in applications and results, but not in their theoretical aspects. In this direction a step forth takes place, to a comparative evaluation of the results in four different methods on: Data Mining, Rough Sets, Neural Networks and Multicriteria Analysis. An original methodological approach of consumer evaluation, combining multicriteria preference-disaggregating analysis and rule induction data mining, are presented **[2]**. **[3]** examined total satisfaction prediction and replacement of missing data of Multicriteria estimations using Data Mining and the Multicriteria Satisfaction Analysis (MUSA) **[1]**. Preference disaggregating methodology is an ordinal regressions based approach **[4]**, **[5]** of Multicriteria Analysis. **[6]** received measurements of customers' satisfaction in coastal shipping using a mathematical model of analytical-synthetical preference modelling.

III. THEORETICAL BACKGROUND

This research examines results, implementing data from 524 questionnaires given to passengers of a big Greek maritime enterprise. Prediction of customers' satisfaction was based on five criteria: Reliability, Prices, Services, Additional Benefits, Comfort-Quality and one total criterion, expressing Total Satisfaction of each passenger. The evaluation scale for these criteria was: Don't Know-Don't Answer (0), Absolute Satisfaction (1), High Satisfaction (2), Regular Satisfaction (3), Less Satisfaction (4), Totally Unsatisfied (5). Those qualitative data were implied elaborating methodologies of: i) Data Mining with Wiz Why system, ii) Rough sets with ROSETTA system, iii) Artificial Neural Networks in MatLab, and iv) Multicriteria Analysis with Multicriteria Hierarchical Discrimination [M.H.DIS] system.

A. Data Mining - Wiz Why

As Data Mining is a decision support process to search for patterns of information in data, based on data retention and distillation. Rule induction models belong to the logical, pattern distillation based approaches of data mining. These technologies extract patterns from data set and use them for various purposes, such as prediction. By automatically exploring the data set, the induction system forms hypotheses that lead to patterns (logic, equations or cross-tabulations). Logic can deal with both numeric and non-numeric data. The central operator in a logical language is usually a variation on the 'if-then' statement. By supervised learning paradigm derive rules, of 'if-then' type, from data. The rule's probability is the probability that for a random record satisfying the rule's condition(s), the rule's conclusion is also fulfilled **[7]**. Rules may easily go beyond attribute-value representations. They may have statements such as 'shipping state = receiving state'. Here in attribute logic, values of the two fields are compared, without naming them. By expressing attribute-based patterns, rules have the advantage of being able to deal with numeric and non-numeric data (categorical fields). Wiz Why employs a sophisticated, non-heuristic algorithm that infers rules from overlapping non-hierarchical sets of conditions, created by the system. Its operation consists of two stages: First, Wiz Why investigates the database considered to be the training set, aiming to reveal all essential regularities that could formulate rules. Next, Wiz Why issues a prediction based on user-defined predicting fields and their values. Wiz Why indicates the probability and significance level for each predicted value of the Field to Predict, which is the dependent variable to be explained and predicted.

B. Rough Sets

ROSETTA system was created by Øhrn, **[8]** in the area of Rough Sets that observe objects which may be indiscernible in terms of descriptors, **[9]**. Rough Sets have been applied in several fields, **[10]**, such as the assessment of firms viability **[11]**. Intuitively a rough set is a set of objects which cannot be precisely characterized in terms of the set of attributes' values. The only sets that can be characterized precisely are lower and upper approximations of objects set. Lower approximation is related to the group of objects that can be classified in the examined set with complete certainty, whilst upper approximation includes objects that belong to the set of objects with a certain probability. Accuracy and quality of approximation can be defined using a lower and an upper approximation of a set, from the interval [0, 1]. Original information required by rough sets, consists of the: I) dependent attribute, II) number of different classes at the dependent attribute, III) independent attributes, IV) number of different classes in each independent attributes, V) training set. An information system, or else known as knowledge representation system in the theory of rough sets, can be viewed as a collection of objects described by values of attributes. Each row in the table represents the information about an object. Formally, an information system S is understood as the 4-tuple:

$$S = U,Q,V,f \quad (2),$$

where U is a finite set of m objects and Q a finite set of n attributes,

$$V = \bigcup q \in Q \; V_q \,(3)$$

and V_q is domain of the attribute q, $f : U \times Q \rightarrow\_V$ is a total function such that f (x, q) $\in V_q$ $\forall q \in Q, x \in U$ can be called an information function. Let $S = U,Q,V$, f be an information systems and let $P \subseteq Q$ and x, y $\in U$. Objects x and y are indiscernible by the set of attributes P in S if:

$$f(x, q) = f(y, q) \; \forall \text{q} \in \text{P} \; (4).$$

Thus every $P \subseteq Q$ generates a binary relation on U, (indiscernibility relation- *IND(P)*). The indiscernibility relation is an equivalence relation for any P. Equivalence classes of *IND(P)* are called P-elementary sets in S and consists of the objects of U that are indiscernible by the set of P. All equivalence classes family of relation *IND(P)* on U is denoted by *U/IND(P)*.

C. Multicriteria Decision Analysis

[12] with M.H.DIS., method proposed an alternative approach for different measures of classification-quality in a complete model deploying additive utility functions that minimize false classifications with clear discrimination in the classes. The previous target is achieved in 2 stages to avoid high computing complexity in MIP problems with many binary variables. M.H.DIS. classifies using of a hierarchical procedure of multicriteria analysis and techniques of mathematical programming. The method classifies defined classes (q- in total) that the first class receives the best items and the last class the worst. Primary stage divides the items of first class (optimal class) form the items of other classes. In the secondary it divides the items of the second from items of lower classes. The process is repeated in q-1 stages to classify all items using 2 additional utility functions for the classification process. Total utility of a decision to classify an item in CR class is UCR(F) and U-CR(F) the non-classification at CR class on stage K. Additional utility functions are provided with mathematical programming. Each stage of the hierarchical process of classification solves 2 problems of linear and 1 of integer programming to detect the best classification to: A) use 2 utility functions that divide items of CR class to items of the worst CR+1,…Cq, to minimize the amount of items false classified to different class, B) adjust of functions maximizing clearance.

D. Neural Networks

Artificial neural networks are learning mechanisms from rather raw electronic models based on the neural structure of the brain **[13]**, which learns from experience creating models, and train them to solve specific problems with behaviour abilities, reaction, shelf-organization, learning, generalizing and oblivion. Even the simplest animal brains are capable to use functions that PC find impossible to implement. The fundamental procedure element is the neuron. Most of the applications demand networks with at least 3 normal types of layers: inputs, hidden and outputs. Inputs layer includes data either from files or directly from electronic sensors. Many hidden layers with multiple neurons in many linked forms exist between the input and output layer. There are 2 types of connections between neurons, that i) accumulate the signal arousing it, and the other causes a deduction, intercepting it, or ii) intercept signals from others neurons in the same layer-posterior interception, used in the input layer. Network chooses the signal with the highest probability and intercepts all the others with posterior interception competition. The next step

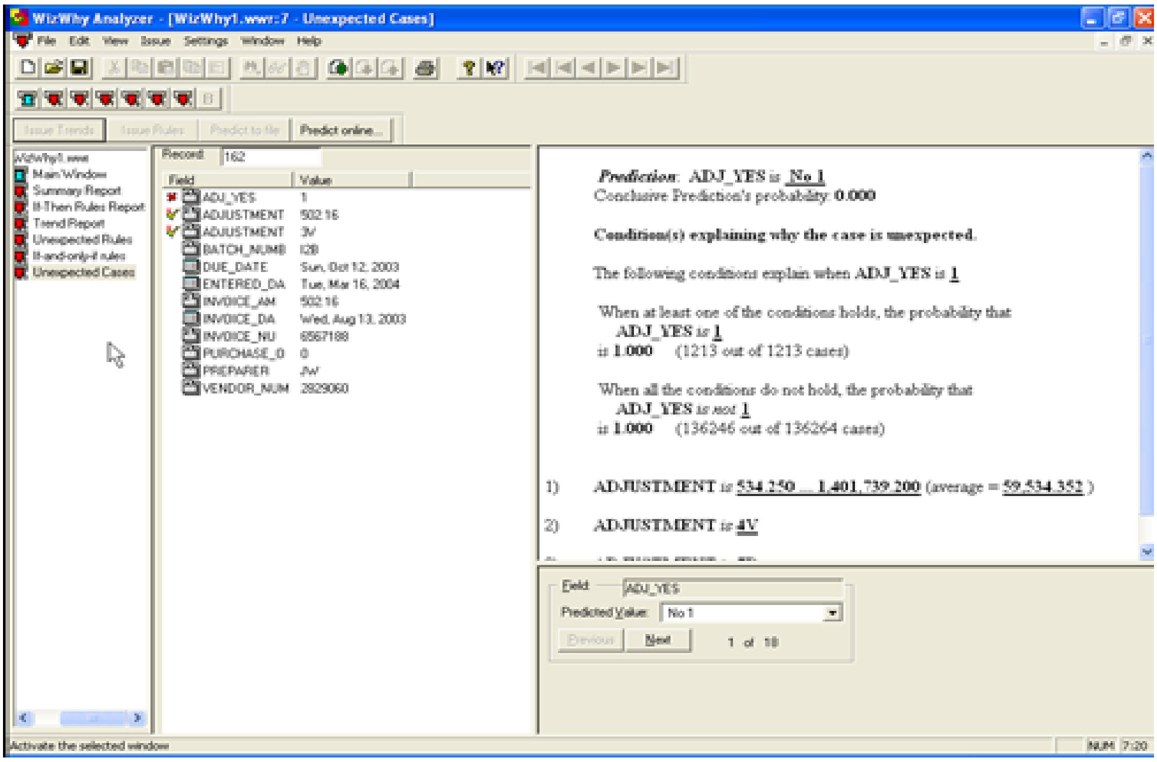

Figure 1. The Wiz Why, during the formation of rules

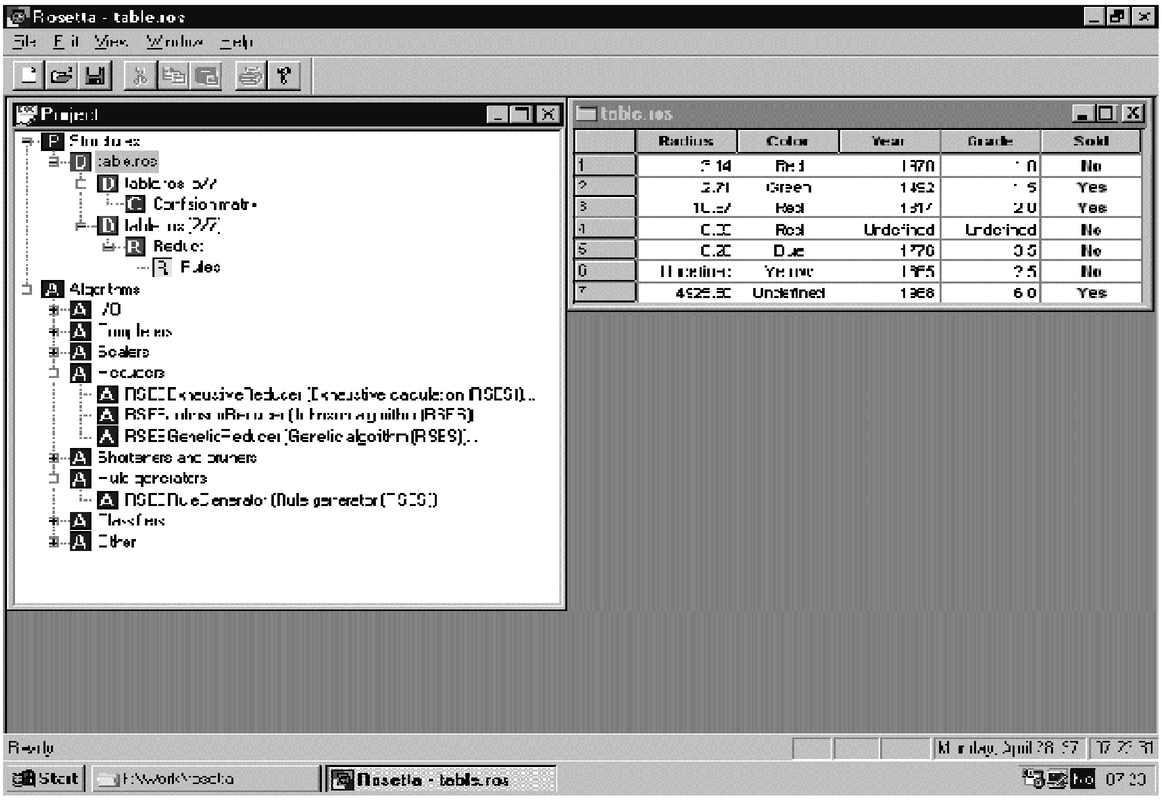

Figure 2. The Rough Sets under ROSETTA software

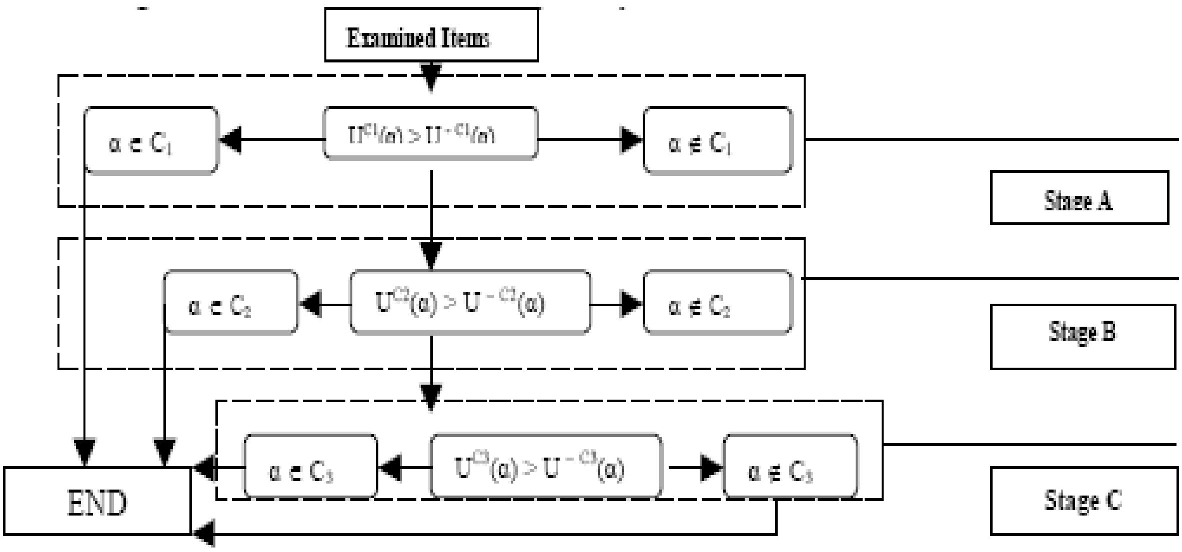

Figure 3. The hierarchical process of classification, M.H.DIS.

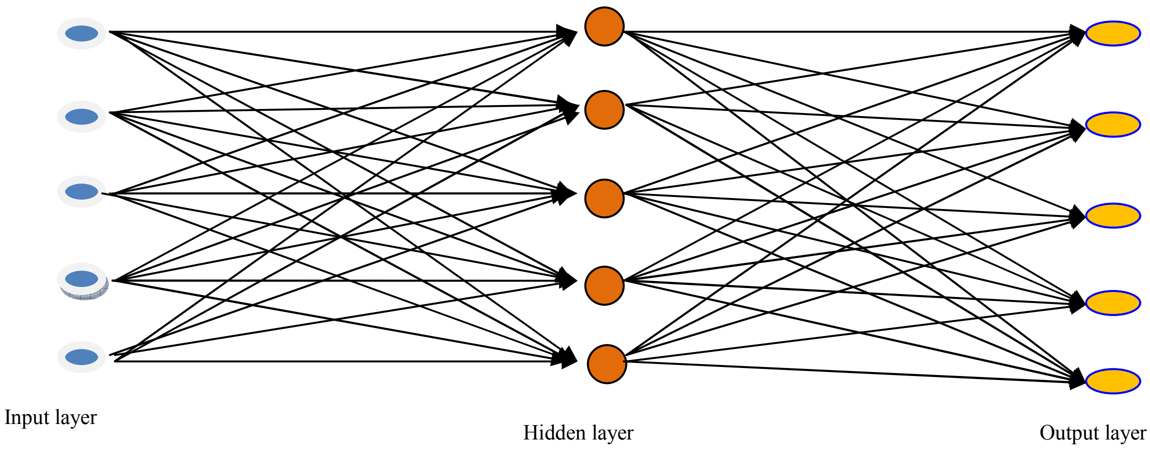

Figure 4. The Artificial Neural Network in the Multi Layer Perceptron topology of 5 neurons per layer

is learning, supervised or unsupervised. **[14]** Loukeris evaluated Radial Basis Functions Networks and hybrids of neuro-genetic RBFNs in Financial Evaluation of Corporations, whilst **[15]** Loukeris et al. revealed a fitter investor performance elaborating exponential utility functions under a variety of heuristics.

IV. Implementation

Primary data have 524 cases: 450 for training and 74 for control. *Wiz Why* **[16]** examines whether values in one field are affected by the values of another. None questionnaire received values 5 or 0 so these were excluded. Optimal solution is varied between a set of 22 to 24 cases per rule, table 1. Solutions gave optimal set of values 11 and 12 cases per rule; solution of 11 cases was chosen as it has maximum number of rules with maximum frequency of percentage. Wiz Why's constrained cases defined frequency of 43.78% whilst the second variable had minimum probability of IF-THEN rules 75% and IF-THEN-NOT at 95%. Results gave 45 rules with probability 1 for the first one that followed declining track until the fifteenth with value 0.763. Correct classification probability of the training set is totally 75%. Correct and error classifications of the control set are on table 2. Probability of correct classification is 71.23% for control set and should be greater than total frequency of the examined.

ROSETTA performs logical analysis of data according to rough sets theory. Training data were imported to the system, and 2 cases were deducted because clients gave zero total satisfaction, thus 448 cases were processed. After the creation of Structures and Algorithms, the number of Algorithm of Holte 1 R, was chosen because rules are created automatically in optimal results, it also provides Singleton' sets of based on Holte theory. ROSETTA rules were deducted without denaturizing the result, with all deduction. It finally produced 25 rules dividing elements in LHS Support (Left Hand Side) and RHS Support (Right Hand Side). A confusion matrix was produced and data of ROC type were returned. On the final stage ROSETTA creates a matrix of correlations classification between predicted training elements and the real ones. Total percentage of correct answers in a sum of 448 valid questionnaires is 64.95%, representing precision of ROSETTA for the 73 control elements. The algorithm Holte 1 R provides optimal results creating rules automatically. ROSETTA produced 21 rules dividing elements in LHS (Left Hand Side) Support and RHS (Right Hand Side) Support. Confusion matrix was produced providing ROC type data. Total percentage of correct answers over a sum of 73 valid questionnaires was 71.23%, which represents accuracy of ROSETTA. Comparing the amount of correct classifications on training set (64.95%) to control data it is concluded that training set had a significantly important difference of 6.28% to control set (71.23%), slightly higher than 5%.

M.H.DIS. on training elements had 119 questionnaires that didn't classify correctly whilst percentage of correct classifications for training set is 73.56%. Training elements matrix gave on the diagonal 71.40% of correct classifications on training data' sum of questionnaires. Analysis gave 16 cases that were classified in different classes so the validity percentage of the method was 78.08% for the control set the classification matrix gave percentage of correct classifications78.34%. Comparing matrices it is concluded that percentage of correct classifications of training set 71.40% has a difference 6.94% from 78.34% of control.

An *Artificial Neural Network* in MATLAB software **[17]**, was elaborated with the feedforward, backpropagation signal transmission and three layers: I) an input layer with five neurons, II) one hidden layer with five neurons and, III) an output layer with five neurons as well. The neural network was in the form of a Multi Layer Perceptron – MLP. The training function was a scaled conjugate gradient backpropagation. Also activation functions of the hidden layer were sigmoid and activation functions for the output layer were linear. Because of problems with zero values on questionnaires (0: Don't know-Don't answer) a new coding took place among 1 to5 (for 0 to 4). A vector of 448 places with values from 2 to 4 represented results. These values are neural networks' estimations for values of passengers' Total Satisfaction. There were some diversifications for the 5 nodes' network, whilst input received 448 training elements and 73 control elements. A vector of 448 elements was the outcome which correlated to the classification of the dependent variable.

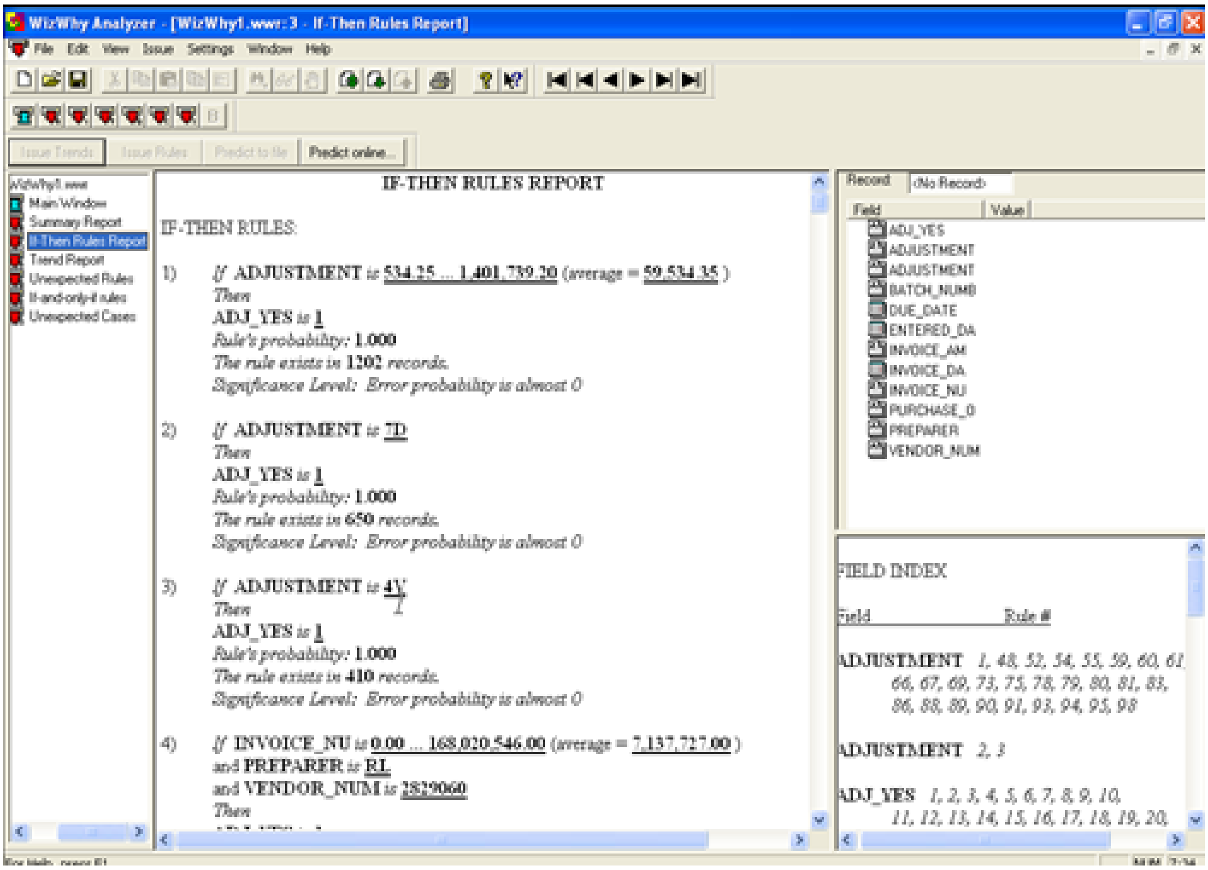

Figure 5a. The Wiz Why, during rules creation

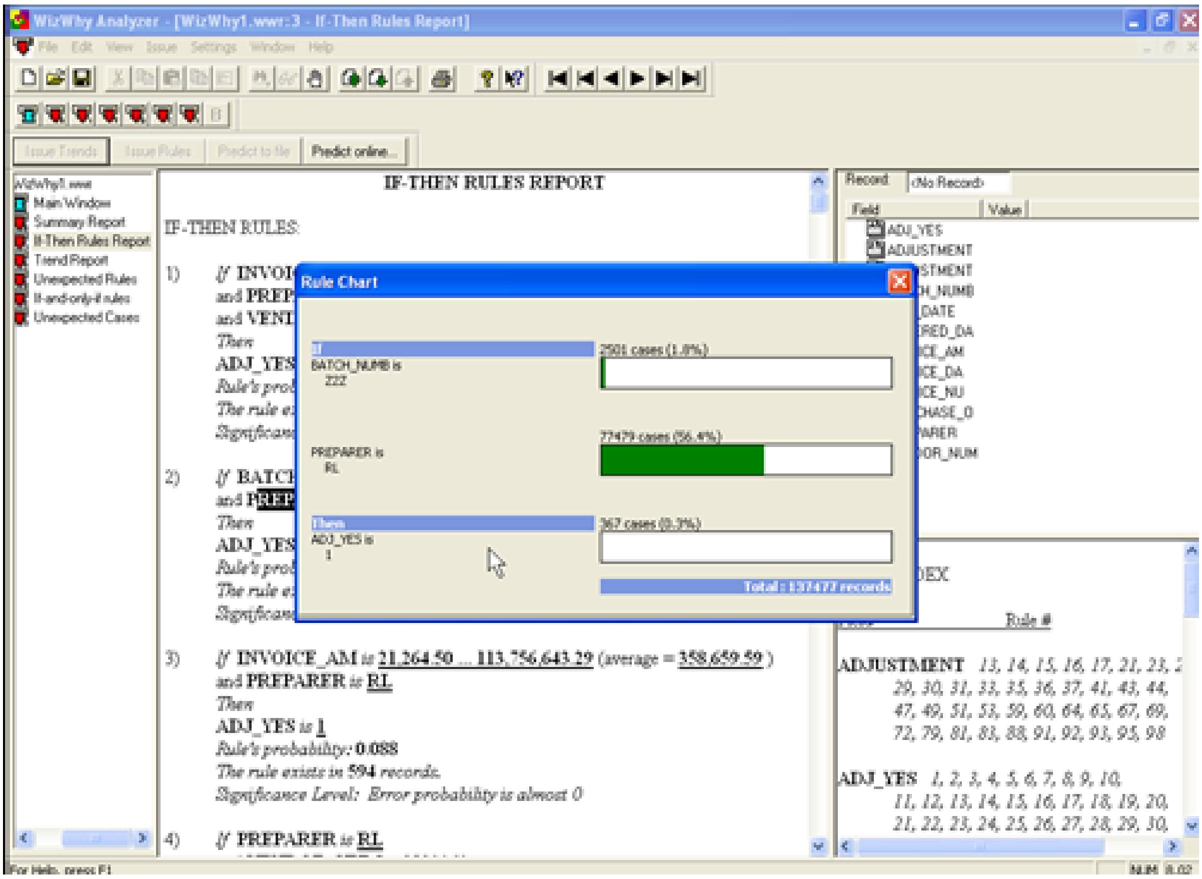

Figure 5b. The Wiz Why, during rules deployment

Table 1: Results in each set of solutions

Number of cases per rule	Number of Rules	Maximum percentage of clients that agrees	Min. percentage of clients that agrees	Frequency of maximum percentage	Frequency of proximate lower percentage' class
10	44	1	0,763	1	5
11	**45**	**1**	**0,763**	**3**	**6**
12	42	1	0,763	3	4
14	38	1	0,763	2	3
16	37	1	0,763	2	3
18	33	1	0,763	2	3
20	31	1	0,763	2	2
22	26	1	0,763	1	1
23	27	1	0,763	1	1
24	27	1	0,763	1	1
30	21	0,892	0,76		
40	14	0,887	0,763		
50	13	0,887	0,763		
60	8	0,851	0,763		
85	3	0,851	0,763		
100	2	0,851	0,791		
120	-	-	-		

Table 2:Confusion matrix of the control set

	1	**2**	**3**	**4**	**Probability**
1	0	0	0	0	0
2	2	27	7	0	0,75
3	1	6	15	1	0,6521
4	0	0	2	10	0,833
					0,7123

150 elements from 448 classified in wrong class, 33.48%, correct classifications had 66.51% (diagonal). The neural network controls autonomously its data, confusion matrix is on table 3.

V. Comparison of Results

The aim of this research is prediction thus control data obtain higher importance. Classification matrices data provided small diversifications in classifications between training and control data (table 4). Most of differences varied in levels of 6.5% (except Wiz Why with the lowest difference of 3.76%) and showed higher percentage of correct classifications for control set whilst training set lacked. This phenomenon was observed on ROSETTA and M.H.DIS. On the contrary Wiz Why methodology antecedences on training set over control set. On the results Wiz Why ranked first in the final ranking of correct classifications of Total Satisfaction per criterion (on training set). M.H.DIS. ranked second with a difference of 3.40% that has higher returns according to Neural Networks in training set and a precedence on correct classifications of control set per 7.11% Networks that ranked on third position and ROSETTA fourth, on control sets and on training sets (M.H.DIS. had 71.40% percentage of success on the training and 78.34% on control, Neural Networks totally 66.51% and ROSETTA 64.95% and 71.23%). Prediction on WizWhy represents optimally the passengers' opinion on the provided services on behalf of the shipping company but also the factors that create the company's image.

VI. Conclusions and Future Research

The possibility to predict customers' satisfaction within a company, using methods of Artificial Intelligence and Multicriteria Analysis was investigated in the current research paper. Each method followed a different philosophy on the elaboration of data, the classification of training and control sets, providing results that determined its efficiency. Data received by customer questionnaires in qualitative form. The important factor in this research prediction ability, thus emphasis is given on classification results of control set. M.H.DIS. had higher success on the control set results from the rest of the three methods, thus it offers better results on the prediction of customers' satisfaction; the method also provided lower percentage on the training set classification with long difference between control and training set classifications. On the other hand data mining (methodology Wiz Why), received the second place in the ranking results of control set classifications but the correct classifications of the training set gave superiority of data mining over the rest methods, with the lowest difference between training and control sets. Rough sets with ROSETTA provided the same results on the control set classifications with the Wiz Why methodology (data mining), offering the lowest results on the training set classification between the methods. Neural networks followed offering total results with a low level of classifications. Concluding the knowledge basis that is created by customers' evaluations is more efficient to be processed firstly with M.H.DIS. or secondly with data mining- Wiz Why methods, to predict future customers satisfaction in high precision.

Table 3: Neural Networks' Classifications

	1	2	3	4	Probabilities
1	8	6	5	0	0,4210
2	10	145	42	6	0,7142
3	1	42	119	24	0,6397
4	2	2	8	26	0,6842
					0,6651

Table 4: Classifications of Methods

METHOD	PERCENTAGE OF TRAINING SET	PERCENTAGE OF CONTROL SET	DIFFERENCE
Wiz Why	75%	71,23%	3,76%
ROSETTA	64,95%	71,23%	6,28%
Neural Nets	66,51%	Total results	
M.H.DIS.	71,40%	78,34%	6,94%

VII. References

[1] Grigoroudis E., & Siskos Y.(2002)-Preference disaggregation for measuring and analysing customer satisfaction: The MUSA method, European Journal of Operational Research

[2] Matsatsinis N.F. (2002)-New Agricultural Products Development using Data Mining Techniques and Multicriteria Methods, (Sideridis Ed.) Proceedings of the 1st Hellenic Assoc. of Information and Communication Technology in Agriculture Food & Environment Conference 6-7/6 Athens

[3] Matsatsinis N., Ioannidou E., & Grigoroudis E. (1999)-Customer satisfaction using data mining techniques, Proc. of European Symposium on Intelligent Techniques ESIT '99, June 3- 4, Hania, Orthodox Academy of Crete

[4] Lagreze J., Siskos J.–Assessing a set of additive utility functions for multicritreria decision making: The UTA method-European Journal of Operat. Research 10.

[5] Siskos Y., & Yannacopoulos D., -UTASTAR : An ordinal regression method building additive value functions-Investigacao Operational, 5 (1) pp 39-53

[6] Grigoroudis E., Malalndrakis J., Politis J., & Siskos Y. (2001), Measuring customer satisfaction in the coastal transportation services, Proceedings of the 12th National

Congress of HELORS, University of Aegean, Pythagoreion Samou, (1), 173-184
[7] Levin B., Meidan A., Cheskis A., Gefen O. & Vorobyov (1999)-PKDD99 Discovery Challenge- Medical Domain, (Berka ed.) PKDD'99 Workshop Notes on Discovery Challenge-Prague pp 55-57
[8] Øhrn Alexander (1999), Discrenibility and Rough Sets in Medicine: Tools and Applications–PhD thesis, Norwegian University of Science and Technology, Dept. of Computer and Information Science, Dec. 1999 NTNU report 1999.
[9] Pawlak Z. (1982)-Rough sets, Int. Jour. of Information and Comp. Sciences, vol. 11, n. 5, 341-356
[10] Slowinski R. (1992)-Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory- Kluwer Academic Publishers, Dordrecht, Netherlands
[11] Slowinski R., Zopounidis C. (1995)-Application of the rough set approach to evaluation of bankruptcy risk, International Journal of Intelligent Systems in Accounting, Finance and Management, vol. 4, pp. 27-41.
[12] Zopounidis C. & Doumpos M. (2000)-Building Additive Utilities for Multigroup Hierarchical Discrimination. The M.H.DIS. method- Optimization Methods and Software 14 (3) 219-240
[13] Sanchez-Sinencio E. & Lau C. (1992), Artificial Neural Networks, IEEE Press.
[14] Loukeris N., Donelly D., Khuman A., Peng Y., (2008), 'A numerical evaluation of meta-heuristic techniques in Portfolio Optimisation', *Operational Research,* Volume 9 (1), ed. Springer Verlang
[15] Loukeris N., (2008), 'Radial Basis Functions Networks to hybrid neuro-genetic RBF Networks in Financial Evaluation of Corporations', *International Journal of Computers*, 2(2)
[16] http://www.wizsoft.com/why.html
[17]http://www.mathworks.com
[18] Loukeris N., Matsatsinis N., (2006), Corporate Financial Evaluation and Bankruptcy Prediction implementing Artificial Intelligence methods, *WSEAS Transactions in Business and Economics*, Issue 4, Volume 3