Rough Sets in insurance sector

M.J. Segovia-Vargas and Z. Díaz-Martínez

Abstract

Rough Set theory methodology belongs to the domain of Artificial Intelligence (AI) and has demonstrated a very high performance in financial issues, especially in classifying problems. Yet, there is little AI research devoted to the insurance industry, although it plays a growing and crucial role in modern economies. The present chapter shows three relevant rough sets researches in insurance sector concluding that this method is an effective tool for supporting managerial decision making in general, and for insurance sector in particular.

1.- INTRODUCCION: ARTIFICIAL INTELLIGENCE AND INSURANCE SECTOR

The Rough Set methodology belongs to the domain of Artificial Intelligence (AI). Artificial intelligence is a new approach in analyzing financial problems. These tools serve as a supplement or complement to statistical methods, but in some cases can act as a substitute for more traditional methods.

Intelligent systems can be constructed in two ways (O'Leary, 1998). The first one is the so-called *Expert Systems*. It consists of introducing knowledge that human experts have accumulated throughout their professional life into a computer. The major limitation to this approach is the process of gathering information because it must be done through a series of interviews with experts. The second approach is the *Machine Learning* one. *Machine Learning* involves developing a computer program capable of generating knowledge through data analysis. This knowledge is used to make inferences about new data. Artificial Neural Networks, Rule Induction Algorithms and Decision Trees are techniques associated with *Machine Learning*. Some of these techniques are explanatory (rule induction and decision trees), while others are characterized by its *black box* nature, such as neural networks.

AI has demonstrated a very high performance in classifying problems. Yet, there is little AI research devoted to the insurance industry, although it plays a growing and crucial role in modern economies. Within the financial sector, the banking one has received more attention from AI researchers. But, the business peculiarities of the insurance sector make impossible to transfer the findings from the banking sector analysis to the insurance one. Therefore a specific analysis is needed (D'Arcy, 2005).

In the past, a large number of methods have been proposed to deal with financial problems in insurance sector. Most approaches used have been statistical techniques such as discriminant or logit analysis (Martín *et al.* 1999, Mora, 1994; Sanchis *et al.*, 2003) and, in many cases, the attributes employed as explicative variables do not satisfy statistical assumptions, which can make difficult to apply them to real problems. Moreover, most real problems consider both qualitative and quantitative factors. This fact can complicate the analysis and the results obtained. Therefore, several classification methods are not suitable when there are qualitative variables. Consequently, in order to avoid limitations of some statistical methods, AI techniques are being applied to tackle financial problems in insurance sector.

Most AI studies devoted to the insurance sector tackle insolvency problems with very satisfactory results (Brocket et al., 1994; Brockett et al., 2006; Díaz, et al. 2005; Kramer, 1997; Martinez de Lejarza, 1996; Salcedo et al., 2004 and 2005; Segovia-Vargas et al. 2004).

As is the case with other methodologies of artificial intelligence, the RS method has been successfully employed to investigate financial problems such as financial distress (Ahn, et al., 2000; Beynon and Peel, 2001; Dimitras, et al 1999; Sanchis, et al., 2007; Slowinski and Zopounidis, 1995; Xiao, *et al*. 2012) activity-based travel modeling (Witlox and Tindemans, 2004), selection of investment projects (Boudreau-Trudel and Kazimierz 2012), investment portfolio and stock analyses (Huang and Jane, 2009; Yao and Herbert, 2009; Shyng *et al.* 2010), e-commerce success indicators (Ahmad, *et al.* 2004) or travel demand analysis (Goh and Law, 2003).

Currently, RS has been applied in the insurance domain (Díaz, *el al.* 2009; Sanchis, et al., 2007; Shyng et al. 2007). The selection of RS method is based not only on its being a high-performance classifying method, but also on its explicative character. This methodology has become a valuable new manner to analyze financial problems since it presents some fundamental advantages, such as the fact that it does not usually need variables to satisfy any assumptions (in contrast with statistical methods); it is possible to use both qualitative and quantitative variables and the elimination of redundant variables is achieved, so the cost of the decision-making process and time employed by the decision- makers are reduced.

In this chapter we present the results of three researches devote to the application of Rough set methodology to the insurance sector.

2.- ROUGH SET THEORY

RS theory was firstly developed by Pawlak (1991) in the 1980s as a mathematical tool to deal with the uncertainty inherent in a decision-making process. Though nowadays this theory has been extended (Greco et al., 1998, 2001), we refer to classical approach. RS theory involves a calculus of partitions; therefore it is related in some aspects to other tools that deal with uncertainty such as statistical probability or fuzzy set theory.

RS approach is somewhat different from either statistical probability or fuzzy set theory. It can be considered that there are three general categories of imprecision in scientific analyses. The first one occurs when events are random in nature; this kind of imprecision is described by statistical probability theory. The second one occurs with objects that may not belong only to one category but may belong to more than one category by differing degrees. In this case the imprecision is associated to the form of fuzziness in set membership and it is the field of fuzzy logic. Finally, RS theory deals with the uncertainty produced when some objects described by the same data or knowledge (so, they are indiscernible) can be classified into different classes (for example, two companies with the same values for some financial variables-they are *indiscernible*- and one of them goes bankrupt and the other one continues in operation) that is, there is not a unique inclusion of these indiscernible objects. This fact prevents their precise assignment to a set. Therefore, the classes in which the objects are to be classified are imprecise, but they can be approximated with precise sets (Nurmi et al., 1996, McKee, 2000).

These differences show one of the main advantages of RS theory: an agent is not required to establish any preliminary or additional information about the data. In the other two categories of imprecision it is necessary to assign precise numerical values to express imprecision of the knowledge, such as probability distributions in statistics or grade of membership or the value of possibility in fuzzy set theory (Pawlak et al. 1995).

The main concept of this approach is based on the assumption that with every object in the universe there can be correlation with associated knowledge and data. Knowledge is regarded in this context as ability to classify objects. Occasionally, objects described by the same data or knowledge are indiscernible in view of such knowledge. The indiscernibility relation leads to the mathematical basis for the RS theory. Intuitively, a RS is a collection of objects that, in general, cannot be precisely characterized in terms of the values of a set of attributes. In real problems or databases, it is usually the occurrence of inconsistencies in classifications. For example, in one of the case of study there are two classes in the database (drivers with and without accident). If a *good* driver (without accidents) has the same attributes as a *bad* one it is difficult to classify them properly into the corresponding classes. To find a solution, there are several ways: the first one consists in increasing the information (for example, considering more attributes or variables) which, sometimes, is not easy or possible. Another possibility is eliminating *these inconsistencies* which is not a proper way because at least some information will be lost. Finally, another way is to deal with these inconsistencies by incorporating them to the analysis (that is RS case).

RS methodology incorporates these inconsistencies creating some approximations to the decision classes. The lower approximation of a class or category consists of all objects that certainly belong to this class and can be certainly classified to this category employing the set of attributes (in our case, the risk factors). The upper approximation of a class contains objects that possibly belong to this class and can be possibly classified to this category using the set of attributes. The difference between the lower and the upper approximation, if it exists, is called the boundary or doubtful region: the set of elements that cannot be certainly classified to a class taking into account the set of attributes. Using the lower and the upper approximation, those classes that cannot be expressed exactly (there is a doubtful region) can be defined precisely using the available attributes.

A fundamental problem of the rough set approach is identifying dependencies between attributes in a database, since it enables the reduction of a set of attributes by removing those that are not essential to characterize knowledge. This problem will be referred as knowledge reduction or, in more general terms, as a feature selection problem. The main concepts related to this question are the core and the reduct. A reduct is the minimal subset of attributes which provides the same classification as the set of all attributes. If there is more than one reduct, the intersection of all of them is called the core and is the collection of the most relevant attributes in the table.

Once the elimination of the redundant variables is achieved, our model can thereafter be developed into the format of the decision rules. Moreover, this technique is explicative and generates decision rules with the following format: "if conditions then decisions". That is, what decisions (actions) should be undertaken when some conditions are satisfied. The number of objects that satisfy the condition part of the rule is called the strength of the rule. The obtained rules do not usually need to be interpreted by an expert as they are easily understandable by the user or decision maker. The most important result in RS approach is the generation of decision rules because they can be used to assign new objects to a class by matching the condition part of one of the decision rule to the description of the object. Therefore, rules can be used for decision support.

3.- ROUGH SET IN INSURANCE SECTOR

3.1 Rough Sets and the prediction of insolvency in insurance sector (Sanchis *et al*. 2007)

3.1.1. The insolvency problem

In the insurance industry, it has long been recognized that there needs to be some form of prudential supervision of such entities to attempt to minimize the risk of failure. Nowadays, Solvency *II* project has led the reform of the existing solvency rules in European Union. Therefore, developing new methods to tackle insolvency problems in insurance sector is a highly topical question.

In general financial terms, insolvency can be referred as the impossibility or inability of a firm to pay its debts and bankruptcy could be interpreted as the culmination of the insolvency process. In this work, the aim is to look for the minimal set of financial ratios that could anticipate possible insolvencies due to permanent financial problems.

Business failure prediction is a classifying problem: firms (objects) described by a set of financial ratios (attributes) are assigned to a category (failed or “healthy” firm).

3.1.2. Analysis and results

Rough set analysis has been performed using ROSE software provided by the Institute of Computing Science of Poznan University of Technology (www-idss.cs.put.poznan.pl/rose. Predki et al., 1998 and Predki and Wilk,1999).

As for the data, it has been employed a sample of Spanish firms used by Sanchis *et al.,* (2003). This data sample consists of non-life insurance firm data five years prior to failure. The firms were in operation or went bankrupt between 1983 and 1994. In each period, 72 firms (36 failed and 36 non-failed) are selected. As a control measure, a failed firm is matched with a non failed one in terms of industry and size (premiums volume). In the analysis, it has been used data one year prior to failure to obtain the decision rules and to test them, it has been used data from years 2, 3, 4 and 5 (Dimitras, *et al.,* 1999).

As for the variables, each firm is described by 17 financial ratios (Table 1).

Table 1: List of Ratios

A1	(Capital+ Reserves)/ Total Liabilities
A5	Working capital/ Total Assets
A6	Current Assets/ Total Assets
B3	Net Premiums/ Total Assets
B6	Provisions for benefit/ Claims Incurred
B7	Net Premiums / (Capital+ Reserves)
B8	(Capital+ Reserves+ Technical provisions) / Earned Premiums
C1	Earnings before Taxes / (Total Liabilities- Capital- Reserves)
C4	Cash-flow / (Capital+ Reserves)
C5	Cash-flow / (Total Liabilities- Capital- Reserves)
C6	Accrued Results / (Subscribed capital- Accrued Results)
C7	Earnings before Taxes / (Capital+ Reserves)
D1	Provisions for benefit/ Earned Premiums
D4	Technical Provisions / Earned Premiums
D6	Technical Provisions / (Capital+ Reserves)
D7	Technical Provisions of cession/ (Capital+ Reserves)
D8	Technical Provisions for current risks/ Earned Premiums

The information table for year 1 which consisted of 72 firms described with 17 ratios and assigned to a decision class (healthy-1-or not-0-) was entered into an input file in ROSE. The financial ratios have been recoded into qualitative terms (low, medium, high and very high) with corresponding numerical values such us 1, 2, 3 and 4 using the quartiles for the values of each variable. This recoding has been made dividing the original domain into subintervals. This recoding is not imposed by the RS theory but it is very useful in order to draw general conclusions from the variables in terms of dependencies, reducts and decision rules (Dimitras *et al.,* 1999).

The first result obtained from RS analysis of the coded information table was that the approximation of the decision classes were equal to one showing that the firms are very well discriminated among them (consequently, the boundary region is empty for the two decision classes).

Another result is that none of the attributes are indispensable for the approximation of the two decision classes (so the core was empty). 452 reducts have been obtained from the table, which contain 4-8 attributes. This result means that, at least, 9 attributes are redundant (and, therefore, they could be eliminated). Consequently, this fact shows the strong support of this approach in feature selection.

It has been selected the reduct consisted of A5, A6, B6, B8, C6, D8 obtaining a reduced table (only six financial ratios) to generate the decision rules. A 30 rule-decision model has been generated and all of them are deterministic.

The rules were tested on data from 2, 3, 4 and 5 years before the actual ratio values (year 1 or year prior to bankruptcy) that were used to obtain the decision rules (Dimitras *et al.,* 1999). The classifications accuracies in percent of correctly classified firms by the set of 30 rules for the five years prior to the reference year (year 1) are shown in Table 2.

Table 2: Rough Sets results

	Year 1	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Rough Set	100%	80.56%	76.36%	75.50%	65.85%

The results are very satisfactory and validate the obtained rules. The rule model shows, from a solvency viewpoint, the importance of these questions: sufficient liquidity, correct rating, proper reinsurance and the need of having enough technical provisions.

3.2 Selection of risk factors in automobile insurance by Rough Sets (Díaz et al. 2009).

3.2.1. Risk factor selection problem

It is well known that insurance companies aim to classify the insured policies into homogeneous tariff classes, assigning the same premium to all the policies belonging to the same class. The classification of the policies into the classes is based on the selection of the so-called *risk factors*, which are characteristics or features of the policies that help the companies to predict their claim amounts in a given period of time (usually one year). In automobile insurance, these are observable variables concerning the driver, the vehicle and the traffic, like age, driving license date, kind of vehicle, circulation zone, etc., that are correlated with the claim rates, and therefore can be useful in order to predict the future claims.

Consequently, it is very important for the insurance company to select an adequate set of risk factors in order to predict the future claim rates correctly and to charge fair premiums to the drivers. The usual approach to select the risk factors is based in statistical multivariate techniques, with mediocre results and leaving a great deal of heterogeneity within the tariff classes. Though, there is a lot of scientific literature dealing with the subject of the risk classification of policyholders. (Denuit *et* al., 2007), during the last years, there have been just a few researches related to AI researches for risk factor selection (Bousoño *et al.,*2008).Therefore, the use of Rough Set theory to tackle this problem could improve the selection of risk factors in automobile insurance to advance in the important issue of premium calculation.

3.2.2. Analysis and results

As for the *data*, it has been used a real sample of 9674 Spanish automobile policies. All data are from 2005. The *risk factors* (variables) employed are 13 and they are both qualitative and quantitative variables. The variables are (Table 3):

Table 3. Variable definition.

Kind of vehicle	This variable takes six values such as car, van, all-terrain vehicle, etc.
Use	Use to which the vehicle is devoted. It takes twenty values: particular, taxi, renting, agrarian use, etc
CV	Power
Private	Private or public vehicle
Tare	Tare (weight)
Plazas	Number of seats of the vehicle
Ambit	Circulation area of the vehicle. This variable takes eight values: international, national, interurban, urban, etc.
Years of the vehicle	The age of the vehicle
Policyholder age	The age of the policyholder.
Driving license	Years of validity of the driving license
Gender	Male or female
Region	Autonomous regions and some big cities such as Valencia, Barcelona and Seville
Diesel	Diesel or gasoline

In this paper rough set analysis has been performed using RSES2 developed by Institute of Mathematics, Warsaw, Poland (http://logic.mimuw.edu.pl/~rses/)

The complete information table contains data from 9674 automobile policies for 2005 described by the 13 variables and assigned to a decision class (accident or not). If a model is developed and tested with the same sample, the results obtained could be conditioned. So in order to avoid it, it has been formed a training set, and a holdout sample to validate the obtained model (decision rules), i.e., the test set. Both sets have been randomly selected. The training information table makes up 70% of total firms and the test information table is made up of the rest of the firms.

The training information table was entered into an input file in RSES2. The continuous variables have been recoded into qualitative terms using subintervals based on the information of the insurance company for all the variables except for the variable Tare (for this variable it has been employed percentiles -10 to 90-).

The first results obtained from RS analysis of the coded information table were the core and the reducts. It has been obtained one reduct containing the whole attributes except PRIVATE. This result means that only this variable is redundant and, therefore, it can be eliminated. Consequently, the company has selected the risk factors properly.

Therefore, through this reduct, a reduced table has been developed to obtain the decision rules. The decision rules have been tested on data from the testing sample, i.e., on the 30% firms that have not been used to estimate the model. The classification accuracy in percent of correctly classified firms is 71%. This result is quite satisfactory and validates both decision rules sets. Moreover, this result is similar to the one obtained by Bousoño *et al. 2008* (70,10%) that have employed other artificial methods for this problem.

The rules show the following preliminary ideas:

- The number of rules for accident class is lower than the number of rules for class no-accident and all rules are deterministic. This means that both classes are well discriminated among them. Therefore the boundary region is empty for the two decision classes. The number of attributes in the rules varies from 5 to 13.

- The use “particular” is associated with accident class.

- The “Ambit” variable shows that “Urban” value appears in all rules that belong to accident class.

- The “Age” variable takes values over forty for all the rules for accident class. If the age is thirty or lower, rules indicate that it is necessary to take into account the age of the driving license to classify policies.

- The variable “region” indicates that driving in some autonomous regions or cities (Seville, Barcelona, Andalucía, Aragón and Cataluña) is associated to no-accident class.

3.3. Rough Set Theory in analyzing the attributes of combination values for the insurance market (Shyng *et al. 2007)*

3.3.1. Problem Description

This research is focus on discovering the customers' need for the insurance market in Taiwan. The fulfillment of customer needs is related to satisfying customer expectations that will provoke they go on buying insurance and/or to repurchase products.

The specialized nature of insurance sector requires up-to-date information to attract potential customers. Moreover, the best way, under highly competitive conditions, to access markets or/and to enlarge market share is acquiring information from these potential customers. One of the main sources of data to get information is designing a questionnaire. In this study a questionnaire with single-choice and multi-choice answers is used to apply Rough Set Theory to investigate the relationship between a single value and a combination of values of attributes.

3.2.2. Analysis and results

Rough Set is used in this paper to analyze the content and the features of data. Rough Set methodology is a very useful tool for exploring data patterns because of its ability to search through a multi-dimensional data space and determine the relative importance of each attribute with respect to its output.

A questionnaire about insurance products has been designed for understanding customer needs for year 2005. This questionnaire has two parts:

- Basic information about participants (that is, participants' personal data). This information will be use to determine purchasing intentions and customers' reasons for not purchasing products.
- A series of questions designed to find information about the purpose of insurance, purchased products, the budget for the premium and reasons for not purchasing insurance products.

As for the *data*, 324 valid replies were obtained from 2005. There are seven attributes:

- Six condition attributes: area, age, gender, profession, purchased products and acceptable premium.
- One decision attribute: Purchase expectation (living or endowment).

The results obtained with Rough Set analysis are very satisfactory because a hit test has been applied to check the feasibility of the decision rules obtaining a 100% hit rate test. Consequently, new data fits the decision classes and the patterns shown by the rules are valid. Therefore, the decision rules show the following customers' insurance needs:

- The purchase purpose is endowment.
- The average annual premium was under US$938
- The targets customers' age 25-35 years
- The most purchased product is a mixture of products.

On the other hand, the given reasons for not purchasing are no interest and the age (too young, below25). If they want to buy products, their purchase expectation is endowment and the premium budget was less than US $ 645.

Finally, the research shows that the main problem was data set classification, because the multi-choice attributes with combination values may enlarge the discrete degree of the data (that is, it increases the number of elementary sets). There are several ways of solving this problem. One is to redefine the attribute values to reduce the range of value classes. This way the number of elementary sets is reduced. To get this redefinition the expert´s contribution in the input data pre-processing task plays an important role.

4.- CONCLUSIONS

Through the exposition we have mentioned some advantages of Rough Set approach so we can conclude that this method is an effective tool for supporting managerial decision making in general, and for insurance sector in particular. In the light of the experiments carried out, it is shown that this theory is a competitive alternative to existing models (statistical or artificial intelligence models) to tackle insurance problems. It has great potential capacities that undoubtedly make it attractive for application to the insurance industry.

Besides empirical results show that this technique offers a great predictive accuracy, it is non-parametric, or distribution free, method, so it does not require the pre-specification of a functional form, or the adoption of restrictive assumptions about the characteristics of statistical distributions of the variables and errors of the model. The decision models are easily understandable and interpretable. This representation of the results makes easier the economical interpretation than other non-parametric techniques. Moreover, the flexibility of the decision rules with changes of the models over the time allows us to adapt them gradually to the appearance of new cases representing changes in the situation.

The lack of Rough Sets research (except for the papers shown) in insurance sector and the satisfactory results obtained by this methodology offer great possibilitiesl for researches in the future.

REFERENCES

Ahmad, F., Hamdan, R. and Bakar, A. (2004). 'Determining Success Indicators of E-Commerce Companies Using Rough Set Approach', *Journal of American Academy of Business, Cambridge* 5, 1/2, 64-69

Ahn, B.S., Cho, S.S. and Kim, C.Y. (2000). 'The Integrated Methodology Rough Set Theory and Artificial Neural Network for Business Failure Prediction', *Expert Systems with Applications,* **18**: 65- 74.

Beynon, M.J. and Peel, M.J. (2001). 'Variable precision rough set theory and data discrimination: An application to corporate failure prediction', *OMEGA: the International Journal of Management Science*, **29**(6): 561–576.

Bousoño, C., Heras, A. and Tolmos, P. (2008). '*Factores de Riesgo y Cálculo de Primas mediante Técnicas de Aprendizaje'*. Editorial MAPFRE, Madrid.

Brockett Patric L., Cooper, W., Golden, L and Pitaktong, U. (1994) 'A Neural Network Method for Obtaining an Early Warning of Insurer Insolvency', *The Journal of Risk and Insurance*, **61**(3): 402-424.

Brockett, P., Golden, L., Jang, J. and Yang.*C*. (2006) ‘A Comparison of Neural Network, Statistical Methods, and Variable Choice for Life Insurers’ Financial Distress Prediction’, *The Journal of Risk and Insurance*, **73**(3): 397-419.

Boudreau-Trudel, B. and Kazimierz, Z. (2012) ‘Comparison of Analytic Hierarchy Process and Dominance-Based Rough Set Approach as Multi-Criteria Decision Aid Methods for the Selection of Investment Projects’ *American Journal of Industrial and Business Management*, 2, 7-12

D’Arcy, S. (2005) *Predictive modeling in automobile insurance: a preliminary analysis. In: World risk and insurance economics congress, August, Salt Lake City*. Working Paper. The Geneva association, No. 302, November.

Denuit, M., Maréchal, X., Pitrebois, S. and Walhin, J.-F. (2007) ‘Index’, in *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*, John Wiley & Sons, Ltd, Chichester, UK

Díaz, Z, Segovia, M.J., Fernández, J. and del Pozo, E. (2005) ‘Machine Learning and Statistical Techniques. An application to the prediction of Insolvency in Spanish Non-Life Insurance Companies’. *The International Journal of Digital Accounting Research*, **5**(9): 1-45.

Diaz, Z. , Tolmos, P. and Segovia M.J. (2009) ´’Selection of Risk Factors in Automobile Insurance by Rough Sets’, XI Congreso-Italiano de Matemática Financiera y Actuarial, Universidad de Badajoz, 2, 3 y 4 de Julio.

Dimitras, A., Slowinski, R., Susmaga, R. and Zopounidis, C. (1999). ‘Business failure prediction using Rough Sets’, *European Journal of Operational Research*, **114**: 263-280.

Goh, C., and Law, R. (2003). ‘Incorporating the rough sets theory into travel demand analysis’. *Tourism Management*, **24**(5): 511–517.

Greco, S., Matarazzo, B., and Slowinski, R. (1998). ‘A new rough set approach to evaluation of bankruptcy risk’, In Zopounidis, C. (ed.), *New Operational Tools in the Management of Financial Risks*, Kluwer Academic Publishers, Dordrecht: 121-136.

Greco, S., Matarazzo, B., and Slowinski, R. (2001). 'Rough Sets Theory for Multicriteria Decision Analysis', *European Journal of Operational Research*, **129**(1): 1-47.

Huang, K-Y and Jane C-J (2009) 'A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories', Expert Systems with Applications 36 5387–5392.

Kramer, B (1997). 'N.E.W.S.: A model for the evaluation of non-life insurance companies'. *European Journal of Operational Research*, **98**(2): 419-430.

Lemaire, J. (1995). *Bonus-Malus Systems in Automobile Insurance*. Boston Kluwer Academic Publisher.

Martín, M.L., Leguey, S. y Sánchez, J. M. Solvencia y estabilidad financiera en la empresa de seguros: Metodología y evaluación empírica mediante análisis multivariante, Cuadernos de la Fundación Mapfre Estudios, vol. 49, 1999.

Martinez de Lejarza Esparducer, I. (1996). 'Forecasting company failure: neural approach versus discriminant analysis. An application to Spanish insurance companies of the 80´s' *Intelligent Systemns in Accounting and Finance,* **5**: 169-186

McKee, T. (2000). 'Developing a Bankruptcy Prediction Model via Rough Sets Theory', *International Journal of Intelligent Systems in Accounting, Finance and Management*, **9**: 159-173.

Mora, A.(1994). 'Los modelos de predicción del fracaso empresarial: una aplicación empírica del logia, Revista Española de Financiación y Contabilidad', 78, enero-marzo, pp. 203-233.

Nurmi, H., Kacprzyk, J. and Fedrizzi, M. (1996). 'Probabilistic, fuzzy and rough concepts in social choice'. *European Journal of Operational Research*, **95**: 264-277.

O'Leary, D. E. (1998). 'Using Neural Networks to Predict Corporate Failure', *International Journal of Intelligent Systems in Accounting Finance and Management*, 7, 187-197.

Pawlak, Z. (1991). *Rough Sets. Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht/ Boston/ London.

Pawlak, Z., Grzymala-Busse, J., Slowinski, R. and Ziarko, W. (1995). Rough Sets.Communicatios of the ACM 38: 11, 89-97.

Predki, B., Slowinski, R., Stefanowski, J., Susmaga, R. and Wilk, S. (1998). ROSE–Software Implementation of the Rough Set Theory. In: L. Polkowski, A. Skowron, eds. Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence, vol. 1424. Springer-Verlag, Berlin, 605-608.

Predki, B. and Wilk, S. (1999). Rough Set Based Data Exploration Using ROSE System. In: Z.W. Ras, A. Skowron (eds.). Foundations of Intelligent Systems, Lecture Notes in Artificial Intelligence, vol. 1609, Springer-Verlag, Berlin, 172-180

Salcedo Sanz, S., Prado Cumplido, M., Segovia Vargas, M.J., Perez Cruz, F. and Bousoño Calzón, C. (2004). 'Feature selection methods involving Support Vector Machines for prediction of insolvency in non-life insurance companies', *Intelligent Systems in Accounting, Finance and Management*, **12**: 261-281.

Salcedo-Sanz, S., Fernández-Villacañas, J. L., Segovia-Vargas M.J. y Bousoño-Calzón, C. (2005). Genetic programming for the prediction of insolvency in non-life insurance companies", Computers and Operations Research, vol.32, nº 4, pp. 749-765, April.

Sanchis, A., Gil, J.A. y Heras, A. (2003). 'El análisis discriminante en la previsión de la insolvencia en las empresas de seguros no vida', Revista Española de Financiación y Contabilidad, 116, enero-marzo, pp.183-233

Sanchís, A., Segovia, M.J., Gil, J.A., Heras, A. and Vilar, J.L. (2007). 'Rough Sets and the Role of Monetary Policy in Financial Stability (Macroeconomic Problem) and the Prediction of Insolvency in the Insurance Sector (Microeconomic Problem)', *European Journal of Operational Research,* **181**: 1554-1573.

Segovia-Vargas M.J., Salcedo-Sanz, S. y Bousoño-Calzón**,** C. (2004) 'Prediction of Insolvency in non-life insurance companies using support vector machines and genetic algorithms', *Fuzzy Economic Review*, IX, 1: 79-94.

Slowinski, R. and Zopounidis, C. (1995). 'Application of the rough set approach to evaluation of bankruptcy risk', *International Journal of Intelligent Systems in Accounting, Finance and Management*, **4**(1):27-41.

Shyng, J-H., Shieh, H-M., Tzeng G-H and Hsieh, S-H (2010) 'Using FSBT technique with Rough Set Theory for personal investment portfolio analysis' European Journal of Operational Research 201, 601-607.

Shyng , J.-Y. ; Wang, F.-K., Tzeng ,G.-H. and Wu, K.-S. (2007). 'Rough Set Theory in analyzing the attributes of combination values for the insurance market', *Expert Systems with Applications*, **32**: 56–64.

Xiao, z., Yang, x., Pang, Y. , and Dang, X. (2012) 'The prediction for listed companies' financial distress by using multiple prediction methods with rough set and Dempster–Shafer evidence theory', Knowledge-Based Systems 26, 196-206.

Yao, J. and Herbert, J. (2009) 'Financial time-series analyses with rough sets', Applied Soft Computing, 9, 1000-1007.

Witlox, F., and Tindemans, H. (2004). 'The application of rough sets analysis in activity-based modeling, opportunities and constraints'. *Expert Systems with Application,* **27**(2): 171–180.