

## Aplicação da Teoria dos *Rough Sets* na Previsão do Fracasso Empresarial

José Manuel Pereira, Fernando Díaz Gómez, Manuel Rodríguez López  
jpereira@ipca.pt, fdiaz@infor.uva.es, marod@udc.es

(recebido em 9 de Maio de 2008; aceite em 3 de Novembro de 2008)

**Resumo:** A previsão do fracasso empresarial é um tema actual que interessa cada vez mais aos diversos agentes económicos, em particular aos investidores, credores, entidades financeiras, mas também aos governos. Na literatura da especialidade existem inúmeros trabalhos que aplicam uma ampla variedade de técnicas para analisar este problema. Neste trabalho apresentamos um método relativamente recente de avaliação do risco de fracasso empresarial com base da teoria dos *Rough Sets*. Esta teoria permite obter um conjunto de atributos de condição que asseguram um nível aceitável de classificação das empresas analisadas, permitindo com isso obter regras para elaborar um sistema de suporte à tomada de decisões. Para a selecção do subconjunto dos atributos relevantes foi utilizado o princípio de comprimento mínimo de descrição aplicado a uma amostra de empresas portuguesas de diversas indústrias. A percentagem global de acertos do modelo estimado para um, dois e três anos antes da ocorrência do fracasso foi de 97,43%, 93,15% e 91,69%, respectivamente.

**Palavras-chave:** Previsão de Insolvência, Fracasso Empresarial; *Rough Sets*, Rácios Financeiros

**Abstract.** Prediction of corporate bankruptcy is a phenomenon of increasing interest to investors, creditors, borrowing firms, and governments alike. In the literature of the specialty we can found a lot of works that apply a wide variety of techniques to analyze this problem. In this work we present a method relatively recent of evaluation of the risk of business failure with base of *Rough Sets* theory. This theory allows obtaining a group of condition attributes that assure an acceptable level of classification of the analyzed companies, allowing with that to obtain rules to elaborate a decision making process system. For the selection of the subset of the relevant attributes was used the minimum description length principle. For the empiric application we have used a sample

of Portuguese companies of mixed industries. The overall predictive accuracy of the model for one, two, and three years before bankruptcy was 97,43%, 93,15% and 91,69%, respectively.

**Keywords:** Bankruptcy Prediction; Financial Distress; Rough Sets; Financial ratios

## 1 Introdução

O problema do fracasso empresarial tem sido, e seguramente continuará a ser, um tema de especial interesse para um amplo conjunto de agentes económicos. O fracasso das empresas pode resultar de um conjunto de causas diversas e complexas, de natureza interna e externa, podendo ser imputado, por exemplo, a uma deficiente estrutura organizativa, à própria estratégia da empresa, modificações tecnológicas ou à evolução da conjuntura económica.

A elaboração de modelos de previsão do fracasso empresarial é um campo de investigação que foi impulsionado pelos trabalhos pioneiros de (Beaver, 1966) e (Altman, 1968) sendo crescente o número de investigadores que se interessam por esta temática.

Na literatura da especialidade existem inúmeros trabalhos que aplicam uma ampla variedade de técnicas para analisar este problema.

Fazendo apenas referência às técnicas multivariantes, podemos destacar como mais utilizadas neste campo de investigação a análise discriminante linear, o logit ou as redes neuronais artificiais. Embora estas técnicas demonstrassem em diversos estudos uma elevada capacidade predictiva, têm como desvantagens o incumprimento de alguns pressupostos metodológicos ou a dificuldade de interpretação das redes neuronais artificiais, que se podem considerar como "caixas negras" que os utilizadores tendem a renunciar pela falta de transparência com que se obtêm os resultados.

Por seu lado, a metodologia dos *Rough Sets* pode proporcionar um conjunto de regras de decisão de fácil interpretabilidade que se obtêm de um conjunto de casos reais do problema objecto de estudo. Cada regra de decisão obtém-se de um conjunto de casos reais, não se requer informação adicional nem é necessário assumir pressupostos sobre os dados.

Esta teoria foi proposta originalmente por Zdzislaw Pawlak nos primeiros anos da década dos 80 (Pawlak, 1982), surgindo da necessidade de dispor de um marco formal para tratar conhecimentos imprecisos, incertos e incompletos expressos na forma de dados adquiridos experimentalmente.

Segundo Díaz (2002) as fontes de incerteza são múltiplas salientando-se a imprecisão do conhecimento disponível, os dados incompletos ou a presença de ruído nos mesmos.

De seguida apresentaremos de forma resumida os principais conceitos e a metodologia de aquisição de conhecimento desta teoria.

Terminaremos este trabalho com uma aplicação empírica da referida teoria a uma amostra de empresas portuguesas de vários sectores de actividade.

## 2. Conceitos fundamentais da Teoria dos Rough Sets

Neste ponto descrevemos os principais conceitos desta teoria, que se poderão ver de forma mais pormenorizada, junto com outras definições, proposições e teoremas em Pawlak (1991).

### 2.1. Sistema de Informação e tabela de decisão

O formalismo utilizado por esta teoria para a representação de dados a analisar reside nos sistemas de informação que se podem representar mediante tabelas atributo-valor (Tabela 1). Formalmente, define-se um sistema de informação como um par  $S = \langle U, A \rangle$ , onde  $U$  representa um conjunto finito e não vazio denominado universo, sendo  $A$  o conjunto de atributos ou características. Cada atributo  $a \in A$  considera-se como uma função total  $a : U \rightarrow V_a$ , donde  $V_a$  é o conjunto de valores possíveis que pode tomar o atributo  $a$ , denominado domínio de  $a$ .

Tabela 1: Tabela atributo-valor

Empresa	Atributos = Variáveis independentes		
	Situação Financeira	Empresa Familiar	Equipa de Gestão
e <sub>1</sub>	Frágil	Sim	Competente
e <sub>2</sub>	Frágil	Sim	Incompetente
e <sub>3</sub>	Frágil	Sim	Muito Incompetente
e <sub>4</sub>	Estável	Sim	Competente
e <sub>5</sub>	Estável	Não	Incompetente
e <sub>6</sub>	Estável	Sim	Muito Incompetente

Fonte: Adaptado de Pawlak *et al.* (1995: 90)

Uma tabela de decisão <sup>1</sup> (Tabela 2)<sup>2</sup> corresponde a uma classe específica de sistemas de informação e formalmente pode-se representar por  $T = \langle U, A, d \rangle$ , onde  $d \notin A$  é o atributo de decisão.

Tabela 2: Tabela de Decisão

Empresa	Atributos = Variáveis independentes			Decisão = Variável Dependente
	Situación Financiera	Empresa Familiar	Equipa de Gestão	Concessão de Crédito
e <sub>1</sub>	Frágil	Sim	Competente	Sim
e <sub>2</sub>	Frágil	Sim	Incompetente	Não
e <sub>3</sub>	Frágil	Sim	Muito Incompetente	Não
e <sub>4</sub>	Estável	Sim	Competente	Sim
e <sub>5</sub>	Estável	Não	Incompetente	Sim
e <sub>6</sub>	Estável	Sim	Muito Incompetente	Não

Fonte: Adaptado de Pawlak *et al.* (1995: 90)

## 2.2. Relação de indiscernibilidade

Um dos principais conceitos na teoria de conjuntos aproximados é a relação de indiscernibilidade (*indiscernibility relation*), a qual normalmente está associada a um conjunto de atributos. Esta relação de indiscernibilidade é uma relação de equivalência. Por exemplo, as empresas e<sub>1</sub> e e<sub>2</sub> apresentam os mesmos valores para os atributos *Situação Financeira* e *Empresa Familiar*. Podemos também verificar que a empresa e<sub>3</sub> é indiscernível das empresas e<sub>1</sub> e e<sub>2</sub> e as empresas e<sub>4</sub> e e<sub>6</sub> são também indiscerníveis entre si.

Pawlak (1991) definiu formalmente a indiscernibilidade da seguinte maneira:

$$IND(B) = \{(x, y) \in U^2 : \forall a \in B, a(x) = a(y)\}$$

Os conjuntos indiscerníveis designam-se de conjuntos elementares. Desta forma, os atributos *Situação Financeira* e *Empresa Familiar* da tabela anterior, definem os seguintes conjuntos elementares:

$$- \{e_1, e_2, e_3\} \rightarrow (Situação\ Financeira=Frágil\ e\ Empresa\ Familiar=Sim)$$

<sup>1</sup> Também designada de tabela de informação.

<sup>2</sup> Exemplo académico. As informações e decisões podem não ter necessariamente correspondência com a realidade.

- $\{e_4, e_6\} \rightarrow (\text{Situação Financeira}=\text{Estável e Empresa Familiar}=\text{Sim})$
- $\{e_5\} \rightarrow (\text{Situação Financeira}=\text{Estável e Empresa Familiar}=\text{Não})$

O conjunto  $\{e_1, e_2, e_3, e_5\}$  é definível através dos atributos *Situação Financeira* e *Empresa Familiar*, pois é possível definir tal conjunto dizendo que qualquer elemento do mesmo é caracterizado pelos atributos *Situação Financeira*=*Frágil* e *Empresa Familiar*=*Sim* ou pelos atributos *Situação Financeira*=*Estável* e *Empresa Familiar*=*Não*. Esta união de conjuntos elementares é designada como conjunto definível.

### 2.3. Redutos e núcleo de conhecimento

Se um conjunto de atributos e o conjunto global definem a mesma relação de indiscernibilidade (quer dizer, os conjuntos elementares das duas relações são idênticos) então, qualquer atributo que pertence ao conjunto global e não pertence ao conjunto de atributos é redundante ou dispensável. No nosso exemplo, o atributo *Empresa Familiar* é redundante.

A noção de reduto (*reduct*) define-se como o mínimo subconjunto de atributos suficientes para manter o mesmo poder de classificação que o conjunto inicial de atributos. Um conjunto de atributos  $P$  que não contém atributos redundantes designa-se como conjunto independente. Um conjunto  $R$  de atributos é um reduto de um outro conjunto  $P$  de atributos, se  $R$  é independente e as relações de indiscernibilidade definidas por  $R$  y  $P$  são as mesmas, isto é,  $IND(R) = IND(P)$ .

Define-se núcleo (*core*) de um conjunto de atributos ao conjunto daqueles atributos que são necessários para manter o mesmo poder de classificação. Formalmente diz-se que um atributo  $a \in P$  é indispensável se  $IND(P) \neq IND(P - \{a\})$  e ao conjunto de todas as características indispensáveis em  $P$  denomina-se núcleo do  $P$ .

Em referência à Tabela 2, o conjunto de atributos  $\{\text{Situação Financeira}, \text{Equipa de Gestão}\}$  é um reduto do conjunto global ou original de atributos  $\{\text{Situação Financeira}, \text{Empresa Familiar}, \text{Equipa de Gestão}\}$ . Eliminando o atributo redundante, temos uma nova tabela de informação (Tabela 3).

Tabela 3: Tabela de Decisão reduzida

Empres a	Atributos = Variáveis independentes		Decisão = Variável Dependente
	Situação Financeira	Equipa de Gestão	Concessão de Crédito
e <sub>1</sub>	Frágil	Competente	Sim
e <sub>2</sub>	Frágil	Incompetente	Não
e <sub>3</sub>	Frágil	Muito Incompetente	Não
e <sub>4</sub>	Estável	Competente	Sim
e <sub>5</sub>	Estável	Incompetente	Sim
e <sub>6</sub>	Estável	Muito Incompetente	Não

Fonte: Adaptado de Pawlak *et al.* (1995: 91)

À semelhança dos conjuntos elementares associados aos atributos, também podemos definir conjuntos elementares associados à decisão, que correspondem aos subconjuntos de todos os exemplos que possuem a mesma decisão. Estes subconjuntos são denominados de conceitos.

Com base na Tabela 3 podemos dizer que um dos conceitos corresponde ao conjunto de empresas a quem foi concedido o crédito (Concessão de crédito = Sim) { e<sub>1</sub> , e<sub>4</sub> , e<sub>5</sub>} e o outro corresponde às empresas a quem foi recusado o crédito (Concessão de crédito = Não) { e<sub>2</sub> , e<sub>3</sub> , e<sub>6</sub>}.

Para exemplificar como é tratada a inconsistência através desta teoria, consideremos a Tabela 4, que resulta da adição das empresas e<sub>7</sub> e e<sub>8</sub> à Tabela 3.

Tabela 4: Tabela de decisão inconsistente

Empres a	Atributos = Variáveis independentes		Decisão = Variável Dependente
	Situação Financeira	Equipa de Gestão	Concessão de Crédito
e <sub>1</sub>	Frágil	Competente	Sim
e <sub>2</sub>	Frágil	Incompetente	Não
e <sub>3</sub>	Frágil	Muito Incompetente	Não
e <sub>4</sub>	Estável	Competente	Sim
e <sub>5</sub>	Estável	Incompetente	Sim
e <sub>6</sub>	Estável	Muito Incompetente	Não
e <sub>7</sub>	Estável	Incompetente	Não
e <sub>8</sub>	Estável	Muito Incompetente	Sim

Fonte: Adaptado de Pawlak *et al.* (1995: 91)

Podemos verificar que os conjuntos elementares da relação de indiscernibilidade definida pelos atributos *Situação Financeira* e *Equipa de Gestão* são {e<sub>1</sub>}, {e<sub>2</sub>}, {e<sub>3</sub>}, {e<sub>4</sub>}, {e<sub>5</sub>, e<sub>7</sub>} e {e<sub>6</sub>, e<sub>8</sub>}, enquanto que os conceitos definidos pela decisão *Concessão de Crédito* são os conjuntos {e<sub>1</sub>, e<sub>4</sub>, e<sub>5</sub>, e<sub>8</sub>} e {e<sub>2</sub>, e<sub>3</sub>, e<sub>6</sub>, e<sub>7</sub>}. Além disso, a decisão *Concessão de Crédito* não depende dos atributos *Situação Financeira* e *Equipa de Gestão*, dado que {e<sub>5</sub>, e<sub>7</sub>} e {e<sub>6</sub>, e<sub>8</sub>} não são subconjuntos de qualquer conceito. Neste caso a Tabela 4 contém informação inconsistente, uma vez que existem empresas com os mesmos valores para os atributos, mas o valor da decisão é diferente.

#### 2.4. Aproximação de conjuntos

Para dirigir a inconsistência, esta teoria recorre ao conceito de aproximação. Para cada conceito  $X$  calcula-se o maior conjunto definível contido em  $X$  e o menor conjunto definível que contém  $X$ , designando o primeiro conjunto de aproximação superior de  $X$ , enquanto que o segundo é denominado de aproximação inferior de  $X$ .

Pawlak (1991) definiu a aproximação inferior, chamada  $\underline{RX}$  como o conjunto de todos os elementos de  $U$  que podem classificar-se, com total certeza, como elementos do conceito de interesse  $X$ , empregando a informação contribuída pelo conjunto de atributos  $R$  y se define formalmente como:

$$\underline{RX} = \bigcup \{Y \in U / R : Y \subseteq X\}$$

A aproximação superior, denominada por  $\overline{RX}$  é o conjunto de todos os elementos de  $U$  que, possivelmente, podem classificar-se como elementos do conceito de interesse  $X$ . Formalmente:

$$\overline{RX} = \bigcup \{Y \in U / R : Y \cap X \neq \emptyset\}$$

Com base na Tabela 4, para o conceito que assume a concessão de um crédito representado pelo conjunto  $\{e_1, e_4, e_5, e_8\}$  a aproximação inferior é definida pelo conjunto  $\{e_1, e_4\}$  e a aproximação superior pelo conjunto  $\{e_1, e_4, e_5, e_6, e_7, e_8\}$  como podemos visualizar a través da Figura 1, elaborada a partir da Tabela 5<sup>3</sup>.

Tabela 5: Tabela de Informação agrupada de tripla entrada

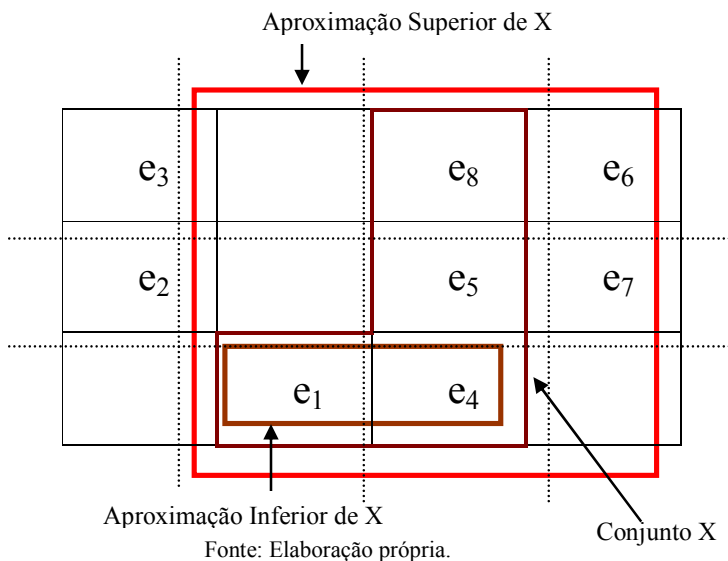
Equipa de Gestão	Concessão de crédito			
	<i>Não</i>	<i>Sim</i>		<i>Não</i>
<i>MI</i>	$e_3$		$e_8$	$e_6$
<i>I</i>	$e_2$		$e_5$	$e_7$
<i>C</i>		$e_1$	$e_4$	
	<i>Frágil</i>		<i>Estável</i>	
	<b>Situação Financeira</b>			

Fonte: Elaboração própria.

<sup>3</sup> Esta tabela de tripla entrada contém a mesma informação que a Tabela 4.



Figura 1: Aproximación superior e inferior de X



Como já referimos, em Pawlak (1991) pode-se ver de forma mais pormenorizada todo o desenvolvimento da referida teoria.

### 3. A Teoria dos *Rough Sets* na Previsão do Fracasso

A teoria dos *Rough Sets* só recentemente se aplicou à previsão da insolvência ou fracasso empresarial. Os primeiros modelos foram todos desenvolvidos na Grécia nos anos noventa (Slowinski & Zopounidis, 1995; Greco *et al.*, 1998<sup>4</sup> e Dimitras *et al.*; 1999). Posteriormente apareceram outros trabalhos, como por exemplo, McKee (2000) ou Rodríguez & Díaz (2005), de que faremos referência a seguir<sup>5</sup>.

<sup>4</sup> Citado por McKee (2000).

<sup>5</sup> McKee & Lensberg (2002) desenvolveram um modelo híbrido utilizando esta teoria e os algoritmos genéticos. Em Tay & Shen (2002) pode-se ver uma síntese de diversos trabalhos que recorreram a esta metodologia.

McKee (2000) utilizou uma amostra com 200 empresas de diferentes sectores (100 fracassadas e 100 não fracassadas) seleccionadas entre os anos 1986 e 1988. No grupo das empresas não fracassadas incluíam-se aquelas que tinham um cash-flow positivo nos últimos 5 anos. As empresas fracassadas foram definidas como aquelas que tinham solicitado a insolvência ou tinham uma filial importante que o tinha feito. Exigia-se que todas as companhias tivessem um mínimo de 5 anos de actividade.

Para o desenvolvimento do modelo o autor utilizou os ratios *Activo circulante/Passivo circulante* y *Resultado neto/Activo total*, obtendo 27 regras, com uma precisão de 93% na amostra de prova e de 88% na amostra de validação.

Rodríguez & Díaz (2005) apresentaram em seu estudo o desenho de um conjunto de modelos de predição do fracasso empresarial utilizando esta metodologia. Para a estimativa dos modelos se utilizou uma amostra constituída por 120 empresas, 60 não fracassadas e 60 fracassadas. Os dados utilizados foram de carácter económico-financeiro (59 ratios) correspondentes a 4 exercícios económicos consecutivos, obtendo um modelo para cada um dos anos prévios ao fracasso. Os resultados obtiveram uma precisão de 91,96% para um ano antes do fracasso, descendendo a precisão à medida que o horizonte temporário aumentava, chegando a 72,14% para o período de 4 anos antes de produzir o fracasso.

Em função dos resultados alcançados, os autores concluem que a teoria de conjuntos aproximados é um método de estimativa que permite obter resultados satisfatórios e modelos facilmente interpretáveis no contexto da previsão do fracasso empresarial.

#### **4. Vantagens e limitações desta teoria**

Según Pawlak *et al.* (1995) uma das principais vantagens dos *Rough Sets* é que não necessitam de informação preliminar ou adicional sobre os dados, como é, por exemplo, a distribuição de probabilidade estatística, já que, como indica Díaz (2002) esta técnica utiliza única e exclusivamente a estrutura interna dos dados analisados para modelar o conhecimento.

McKee (2000) faz referência também, a que as regras de classificação são de fácil interpretação e que cada regra de decisão se obtém de um conjunto de casos reais.

Bose (2006) salienta que através das regras descobertas é possível obter uma explicação da decisão.

Para Santos & Azevedo (2005) o fato de esta teoria se fundamentar num rigoroso formalismo matemático é também uma vantagem. Como limitações estes autores

mencionam a dificuldade de execução e também a documentação ainda pobre e complexa.

## **5. Aplicação empírica**

### **5.1. Definição de fracasso utilizada**

Nos estudos sobre a previsão do fracasso empresarial geralmente as empresas dividem-se em duas categorias: fracassadas e não fracassadas.

No nosso estudo identificamos uma empresa como fracassada se tiver ocorrido alguma das seguintes situações:

- processo de insolvência ou a mesma tenha sido declarada pelos tribunais;
- processo de reestruturação;
- tenha pendente processos judiciais por reclamação de dívidas.

A referida informação foi obtida através da III Série do Diário da República, do Instituto Informador Comercial e junto de Gestores Judiciais.

### **5.2. Selecção das variáveis**

A falta de uma base teórica que explique o fracasso empresarial e sirva de orientação no processo de selecção das variáveis explicativas permitiu que se utilizassem múltiplos critérios na selecção das mesmas.

À semelhança do que sucedeu com diversos estudos a selecção das variáveis independentes – no presente estudo, rácios económico-financeiros – teve por base a sua maior utilização e nível de significância obtido em diversos estudos publicados na literatura sobre o tema.

A lista dos 60 rácios seleccionados que se elaboraram com a informação do Balanço e da Demonstração dos Resultados das empresas que compõem a amostra encontra-se no Anexo I.

### **5.3. Procedimento de selecção da amostra**

A informação para elaborar a amostra de empresas obteve-se da base de dados SABE da empresa Informa, S.A (dados disponíveis na Biblioteca da Universidade de Vigo até 31 de Dezembro de 2003), da Dun & Bradstreet, do Instituto Informador Comercial e de Gestores Judiciais.

Para fazerem parte da amostra as empresas tinham que apresentar dados para 3 exercícios económicos consecutivos, compreendidos entre o período de 2000 e 2002. Seleccionou-se uma amostra composta por empresas pertencentes a diversos sectores de actividade, com sede em Portugal continental, incluindo indústrias do

sector têxtil, vestuário, confecção, madeira, cortiça, calçado, couro, plásticos, moldes, vidro, cutelaria e cerâmicas.

A amostra era constituída por 818 empresas, sendo 742 consideradas saudáveis ou não fracassadas e 76 fracassadas.

#### **5.4 Análise e Resultados**

Para procedermos à análise empírica aplicando a teoria dos *Rough Sets* utilizamos um software desenvolvido por Díaz (2000).

Uma das operações mais importantes na fase de pré-processamento é efectuar a discretização dos valores dos atributos (rácios). Estes foram discretizados em 5 categorias com igual frequência (*equal frequency binning*), cujo resultado para os valores dos ratios das empresas que compõem a amostra se pode ver na Tabela 6<sup>6</sup>.

---

<sup>6</sup> Por uma questão de economia de espaço fazemos apenas referência aos intervalos de discretização para os 10 primeiros atributos. Os dados completos serão disponibilizados quando solicitados aos autores.

Tabela 6: Discretização dos atributos

- Atributo	- 0	- 1	- 2	- 3	- 4
- X1	- ]*, 1.18144 [	- [1.18144, 1.36419 [	- [1.36419, 1.53683 [	- [1.53683, 1.85887[	- [1.85887, *[[
- X2	- ]*, 0.47699[	- [0.47699, 0.69996[	- [0.69996, 0.92569[	- [0.92569, 1.34922[	- [1.34922, *[[
- X3	- ]*, - 0.09372[	- [- 0.09372, 0.10827[	- [0.10827, 0.29041[	- [0.29041, 0.57611[	- [0.57611, *[[
- X4	- ]*, - 0.05754[	- [- 0.05754, 0.05315[	- [0.05315, 0.13964[	- [0.13964, 0.30047[	- [0.30047, *[[
- X5	- ]*, 0.41664[	- [0.41664, 0.57106[	- [0.57106, 0.68506[	- [0.68506, 0.82800[	- [0.82800, *[[
- X6	- ]*, 0.86883[	- [0.86883, 1.14998[	- [1.14998, 1.42226[	- [1.42226, 1.99607[	- [1.99607, *[[
- X7	- ]*, 0.58187[	- [0.58187, 0.80674[	- [0.80674, 1.01154[	- [1.01154, 1.28636[	- [1.28636, *[[
- X8	- ]*, 0.20818[	- [0.20818, 0.46092[	- [0.46092, 0.75194[	- [0.75194, 1.40584[	- [1.40584, *[[
- X9	- ]*, 0.02052[	- [0.02052, 0.03634[	- [0.03634, 0.05882[	- [0.05882, 0.10230[	- [0.10230, *[[
- X10	- ]*, 0.15333[	- [0.15333, 0.26622[	- [0.26622, 0.34814[	- [0.34814, 0.46188[	- [0.46188, *[[

Fonte: Elaboração própria

O algoritmo de selecção dos atributos baseia-se no princípio MDL (*Minimum Description Length*). Com este princípio, em cada passo do algoritmo elimina-se aquela variável que, simultaneamente, maximiza o ganho verificado pela diminuição da complexidade do modelo resultante e minimiza a perda de precisão que sofre o mesmo. (Díaz, 2001).

Na fase do processamento a amostra foi dividida em duas, uma de treino e outra de validação. A amostra de treino estava constituída por 546 empresas (495 não fracassadas e 51 fracassadas) correspondendo a dois terços das empresas, tendo como referência o ano prévio ao fracasso, sendo a amostra de validação constituída pelas restantes empresas (247 não fracassadas e 25 fracassadas).

O modelo foi também validado com os dados de todas as empresas referentes a dois e três anos antes da ocorrência do fracasso.

O método de inferência utilizado tinha como critério escolher a regra satisfeita por uma nova observação, com o objectivo de ter o maior número de observações dentro do conjunto de treino, suportando-a, isto é, que a satisfaça.

Na fase de treino o erro máximo admitido era de 2,5% (beta = 0.025), tendo-se obtido 121 redutos com base no seguinte conjunto P de atributos:

$$P = \{ X4, X5, X8, X20, X31, X35, X39, X44, X47, X51, X55 \}$$

O reduto número 23 foi o que apresentou melhor comportamento global, sendo obtidas 14 regras geradas a partir de  $\{ 20,47,51 \}$ <sup>7</sup>, cujas primeiras três descrevemos a seguir<sup>8</sup>.

$$r\_0: [ (X20 = 4) ] \Rightarrow [ (Decisão = 1) ]$$

CoverP={0,2,3,5,10,13,15,27,28,32,34,35,37,44,48,56,57,59,66,67,69,70,71,81,82,90,91,93,98,101,104,105,110,112,128,134,135,139,147,148,149,151,152,158,159,161,173,175,176,177,179,180,183,184,187,189,190,191,195,196,206,207,212,214,217,218,219,223,230,231,234,235,236,242,246,247,253,255,256,263,266,267,272,276,279,283,284,286,291,293,295,298,302,303,306,332,336,338,352,360,364,365,368,369,370,372,375,378,380,381,387,388,390,391,394,401,410,412,427,432,447,448,451,455,464,468,471,476,479,481,490,494,515,524 } Erro= 0.0149

$$r\_1: [ (X51 = 4) ] \Rightarrow [ (Decisão = 1) ]$$

CoverP={1,5,10,16,25,28,29,30,32,34,35,37,38,41,44,45,51,56,59,64,65,66,68,69,70,71,74,81,84,88,91,92,93,101,103,104,105,109,110,111,116,123,135,144,146,147,148,152,153,154,156,158,159,161,163,168,170,177,179,183,185,187,191,192,193,197,200,203,206,207,216,217,222,223,224,225,229,230,233,234,247,248,251,256,267,268,271,272,273,275,282,286,288,291,292,298,300,302,303,309,317,318,326,331,334,338,340,341,349,350,351,352,354,358,362,364,365,370,373,378,380,382,385,387,388,390,391,394,397,399,401,410,412,418,419,421,422,423,426,427,429,430,441,451,452,457,462,464,468,469,470,476,482,493,494 } Erro= 0.0000

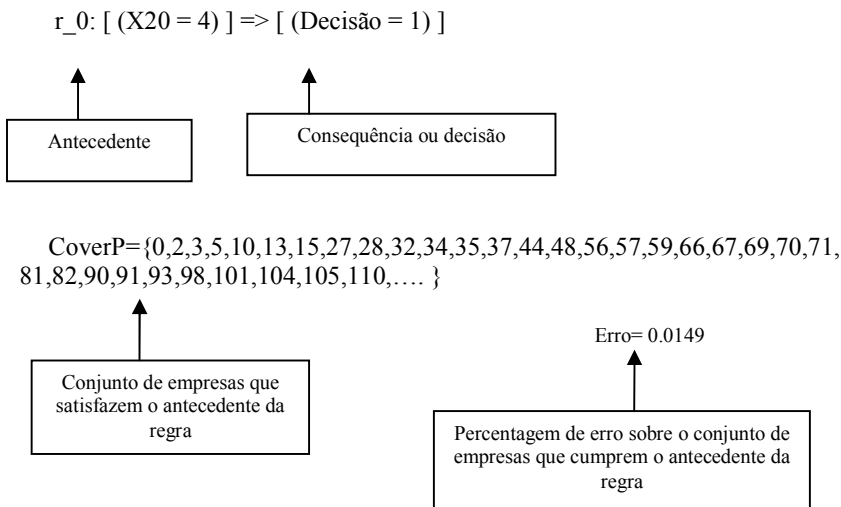
$$r\_2: [ (X51 = 3) ] \Rightarrow [ (Decisão = 1) ]$$

CoverP={2,6,17,22,36,40,48,50,55,60,73,83,85,94,98,99,102,106,108,112,113,114,117,118,120,121,124,126,127,129,131,132,136,137,143,145,149,151,155,162,165,166,169,176,178,182,184,186,189,194,195,202,204,205,209,210,212,215,227,231,235,236,245,260,261,263,264,266,269,270,278,279,280,283,290,293,294,295,297,301,304,311,314,315,319,322,324,332,335,345,346,348,357,359,361,363,368,369,371,372,374,375,376,379,383,389,395,402,404,405,407,431,432,433,436,437,438,440,442,445,446,447,448,454,458,459,465,466,467,472,473,475,487,488,491 } Erro= 0.0000

<sup>7</sup> X20, X47 e X51 correspondem aos seguintes indicadores: (Disponibilidades + Créditos a Curto Prazo) / Passivo a Curto Prazo, Resultado líquido/Capital Próprio e Resultados Operacionais/Activo Total, respectivamente.

<sup>8</sup> As restantes encontram-se no Anexo 2.

Para exemplificar a interpretação de cada uma das regras, recorreremos à primeira, que tem o código  $r_0$ :



Os resultados obtidos pelo modelo aplicado a todas as amostras de empresas encontram-se nas tabelas seguintes, estando as empresas não fracassadas representadas por “1” e as fracassadas por “0”.

Como já havíamos referido a amostra de treino estava formada por 2/3 dos dados disponíveis referentes a um ano antes de ocorrer o fracasso. Com a referida amostra o modelo obteve 540 decisões correctas de um total de 546, com um erro global de 1,10% (Tabela 7).

Tabela 7: Tabela de contingência (amostra de treino)

<b>Valor Esperado</b>	<b>Valor Inferido</b>	
	0	1
0	45	6
1	0	495

A amostra de validação referente a um ano antes do fracasso era composta por 272 empresas, 25 fracassadas e 247 não fracassadas. Das 272 empresas o modelo acertou em 265, registando um erro global de 2,57%.

Tabela 8: Tabela de contingência (amostra de validação)

<b>Valor Esperado</b>	<b>Valor Inferido</b>	
	0	1
0	20	5
1	2	245

Para os dados referentes a 2 anos antes do fracasso o modelo manteve uma percentagem global de acertos elevada (6,85%), decidindo acertadamente para 762 empresas das 818 que faziam parte da amostra. No entanto, o erro tipo I, (que corresponde ao facto de classificar uma empresa fracassada como saudável) é relativamente elevado, superior a 60%.

Tabela 9: Tabela de contingência (amostra de validação para 2 anos prévios)

<b>Valor Esperado</b>	<b>Valor Inferido</b>	
	0	1
0	29	47
1	9	733



O comportamento do modelo quando aplicado à amostra referente aos dados para 3 anos antes da ocorrência do fracasso foi semelhante ao registado pela amostra anterior. Podemos verificar que em termos globais o modelo regista resultados muito satisfatórios, classificando correctamente 750 das 818 empresas, a que corresponde uma percentagem de acertos de 91,69%. O erro tipo II (que corresponde ao facto de classificar uma empresa não fracassada como fracassada) manteve-se inferior a 2%, continuando o erro tipo I muito elevado, como se pode comprovar na Tabela 10.

Tabela 10: Tabela de contingência (amostra de validação para 3 anos prévios)

<b>Valor Esperado</b>	<b>Valor Inferido</b>	
	0	1
0	20	56
1	12	730

## Conclusão

Neste trabalho elaboramos um modelo de previsão do fracasso empresarial utilizando a teoria dos *Rough Sets*, recorrendo a uma amostra de empresas portuguesas de diferentes indústrias.

As regras do reduto que apresentou melhor comportamento global foram geradas a partir dos seguintes indicadores: (*Disponibilidades + Créditos a Curto Prazo*) / *Passivo a Curto Prazo*, *Resultado liquido/Capital Próprio* e *Resultados Operacionais/Activo Total*.

A capacidade predictiva do modelo com as amostras de treino e validação referentes a um ano antes do fracasso foi de 98,9% e 97,4%, respectivamente, diminuindo a sua precisão à medida que nos distanciamos no horizonte temporal, registando uma percentagem de acertos de 91,69% com os dados referentes a três anos antes.

Dos resultados obtidos podemos também concluir que o erro tipo I (que corresponde ao facto de classificar uma empresa fracassada como saudável) é elevado, principalmente nas amostras referentes a dois e três anos antes do fracasso. Pensamos que o facto de mais de 80% das empresas da amostra não terem as suas contas auditadas, (não estavam a isso obrigadas) as suas contas podem não reflectir a real situação financeira e económica, nomeadamente nas empresas fracassadas, a estimação do modelo com amostras relativamente desequilibradas no que se refere

ao número de empresas em cada um dos grupos bem como a utilização de empresas de diferentes indústrias pode ter influenciado no comportamento do modelo, principalmente no segundo e terceiro ano antes do fracasso.

Com base nas percentagens de acerto globais podemos concluir que a teoria dos *Rough Sets* pode proporcionar modelos de previsão do fracasso empresarial com resultados muito satisfatórios baseados em regras facilmente interpretáveis.

Como extensão deste trabalho pretendemos realizar uma análise comparativa desta metodologia com outras técnicas e modelos estatísticos e de inteligência artificial que têm sido utilizados neste campo de investigação, utilizando uma amostra de empresa de um único sector de actividade.

## Bibliografía

- Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), 589-609.
- Beaver, W. H. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research*, 4, 71-111.
- Bose, I. (2006). Deciding the financial health of dot-coms using rough sets. *Information & Management*, 43(7), October, 835-846.
- Díaz Gómez, F. (2001). Aprendizaje y Generación Automática de Conocimiento: Construcción de Redes Bayesianas mediante Rough Sets. *Revista Iberoamericana de Inteligencia Artificial*, 20, 81-84.
- Díaz Gómez, F. (2002). Aprendizaje y Generación Automática de Conocimiento: Construcción de Redes Bayesianas mediante Rough Sets. *Tesis Doctoral*, Universidad de Vigo, Vigo.
- Dimitras, A. I.; Slowinski, R.; Susmaga, R. & Zopounidis, C. (1999). Business Failure Prediction Using Rough Sets. *European Journal of Operational Research*, 114(2), April, 263-280.
- Greco, S.; Matarazzo, B. & Slowinski, R. (1998). A New Rough Set Approach to Evaluation of Bankruptcy Risk. In Zopounidis (Ed.), *Operational Tools in the Management of Financial Risk*, (pp. 121-136). Kluwer: Dordrecht.
- McKee, T. E. (2000). Developing a Bankruptcy Prediction Model via Rough Sets Theory. *Internacional Journal of Intelligent Systems in Accounting, Finance & Management*, 9(3) September, 159-173.
- McKee, T.E. & LENSBERG, T. (2002). Genetic programming and rough sets: A hybrid approach to bankruptcy classification. *European Journal of Operational Research*, 138 ( 2), April, 436-451.
- Pawlak, Z. (1982). Rough Sets. *International Journal of Computer and Information Sciences*, 11, 341-356.
- Pawlak, Z. (1991). *Rough Sets: Theoretical Aspects of Reasoning about Data.*, Kluwer Academic Publishers, Dordrecht.
- Pawlak, Z.; Grzymala-Busse, J.; Slowinski, R. & Ziarczo, W. (1995). Rough Sets. *Communications of the ACM*, 38 (11), 89-95.
- Rodríguez López, M. & Díaz Gómez, F. (2005). La Teoría de los “Rough Sets” y la Predicción del Fracaso Empresarial. Diseño de un Modelo para las PYMES. *XIII Congreso AECA*, septiembre, Oviedo.
- Santos, M. F. & Azevedo, C. (2005). *Data Mining – Descoberta de Conhecimento em Base de Dados*. FCA: Lisboa.
- Slowinski, R. & Zopounidis, C. (1995). Application of the Rough Set Approach to Evaluation of Bankruptcy Risk. *Intelligent Systems in Accounting, Finance and Management*, 4(1), March, 27-41.
- Tay, F. E. H. & Shen, L. (2002). Economic and financial prediction using rough sets model. *European Journal of Operational Research*, 141(3), September, 641-659.

## Anexo 1: Rácios Seleccionados

X1	Activo Total / Passivo Total
X2	(Activo Circulante - Existências) / Passivo Circulante
X3	(Activo Circulante - Passivo Circulante) / Passivo total
X4	(Activo Circulante - Passivo Circulante) / Vendas
X5	Activo Circulante / Activo Total
X6	Activo Circulante / Passivo a Curto Prazo
X7	Activo Circulante / Passivo Total
X8	Activo Fixo / Activo Circulante
X9	Amortizações / Proveitos Operacionais
X10	Capital Próprio / Activo Total
X11	Capital Próprio / Imobilizado Liquido
X12	Capital Próprio / Passivo Total
X13	Cash-flow / Passivo a Curto Prazo
X14	Cash-flow / Passivo Total
X15	Custos Financeiros / Proveitos Operacionais
X16	Custos Financeiros / Resultados Operacionais
X17	Custos Financeiros / Total Dividas a Instituições de Crédito
X18	Disponibilidades / Activo Total
X19	Disponibilidades / Passivo a Curto Prazo
X20	(Disponibilidades + Créditos a Curto Prazo) / Passivo a Curto Prazo
X21	Dividas a terceiros / Activo Total
X22	Existências / Activo Total
X23	Existências / Vendas
X24	Capital circulante / Activo Total
X25	Custos com Pessoal / Activo Fixo
X26	Custos com Pessoal / Proveitos Operacionais
X27	Imobilizado Fixo/ Activo Circulante
X28	Imobilizado Fixo / Activo Total
X29	Imobilizado Intangível / Activo Total
X30	Imobilizado Liquido/ Activo Total
X31	Imobilizado Total / Dividas a Longo Prazo

X32	$(\text{Investimentos Fin. M/L Prazo} + \text{Aplicações Fin. C/P}) / \text{Activo Total}$
X33	$(\text{Aplicações Fin. C/P} + \text{Disponibilidades}) / \text{Activo Total}$
X34	$\text{Passivo a Curto Prazo} / \text{Passivo Total}$
X35	$\text{Passivo a Médio e Longo Prazo} / \text{Passivo a Curto Prazo}$
X36	$\text{Passivo a Médio e Longo Prazo} / \text{Passivo Total}$
X37	$\text{Passivo Total} / \text{Capital Próprio}$
X38	$\text{Proveitos Operacionais} / \text{Activo Circulante}$
X39	$\text{Proveitos Operacionais} / \text{Activo Total}$
X40	$\text{Proveitos Operacionais} / \text{Custos Operacionais}$
X41	$\text{Proveitos Operacionais} / \text{Imobilizado}$
X42	$(\text{RAI} + \text{Amort.} + \text{Provisões}) / \text{Custos Financeiros}$
X43	$\text{Reservas} / \text{Activo Total}$
X44	$(\text{Reservas} + \text{Resultados Exerc. Anteriores}) / \text{Passivo Total}$
X45	$\text{Resultados Extraordinários} / \text{Resultado líquido}$
X46	$\text{Resultado líquido} / \text{Activo Total}$
X47	$\text{Resultado líquido} / \text{Capital Próprio}$
X48	$\text{Resultado líquido} / \text{Passivo Total}$
X49	$\text{Resultado líquido} / \text{Proveitos Operacionais}$
X50	$\text{Resultado líquido} / \text{Vendas}$
X51	$\text{Resultados Operacionais} / \text{Activo Total}$
X52	$\text{Resultados Operacionais} / \text{Custos Financeiros}$
X53	$\text{Resultados Operacionais} / \text{Proveitos Operacionais}$
X54	$\text{Resultados Operacionais} / \text{Vendas}$
X55	$\text{Vendas} / \text{Activo Total}$
X56	$\text{Vendas} / \text{Capital Circulante}$
X57	$\text{Vendas} / \text{Disponibilidades}$
X58	$\text{Vendas} / \text{Existências}$
X59	$(\text{RAI} + \text{Custos Financeiros}) / \text{Vendas}$
X60	$\text{RAI} / (\text{RAI} + \text{Custos Financeiros})$

## Anexo 2: Restantes Regras do Reduto Seleccionado

$r\_3: [ (X51 = 2) ] \Rightarrow [ (Decisão = 1) ]$

CoverP={3,4,13,19,20,21,24,26,31,43,46,49,52,53,57,61,62,67,72,75,76,77,79,80,86,90,95,97,100,107,122,130,138,139,140,141,142,150,160,167,174,180,181,190,196,199,201,211,214,218,220,226,228,232,240,242,244,246,250,254,259,262,265,276,284,287,299,305,306,310,312,316,320,330,333,336,337,339,343,344,347,353,356,360,366,367,381,392,398,400,403,406,408,409,413,414,415,416,417,420,424,428,434,435,439,443,444,455,460,461,471,474,477,478,479,480,484,486,489,490 }  
Erro= 0.0000

$r\_4: [ (X47 = 3) ] \Rightarrow [ (Decisão = 1) ]$

CoverP={6,13,21,22,26,32,34,37,43,45,50,51,52,58,59,60,61,64,65,66,68,70,71,74,77,80,85,92,93,102,104,106,110,111,112,114,116,118,119,121,131,135,142,144,150,151,155,162,163,168,169,176,178,180,185,191,192,195,196,200,205,212,216,218,225,230,231,232,234,235,240,241,247,251,256,261,266,267,275,278,279,280,281,291,296,298,315,320,324,325,327,331,341,345,346,349,350,354,356,357,363,366,368,369,370,372,375,378,379,380,381,383,385,390,394,397,398,404,407,411,416,419,421,422,423,425,427,430,433,436,440,441,443,444,445,446,447,451,454,457,460,464,466,467,472,477,479,480,482,485,488 } Erro= 0.0000

$r\_5: [ (X47 = 1) ] \Rightarrow [ (Decisão = 1) ]$

CoverP={7,9,11,15,18,23,27,33,42,47,49,53,54,57,63,72,75,76,82,87,89,96,97,115,120,127,128,140,141,143,154,157,164,165,166,167,171,174,181,190,197,198,199,202,209,213,219,226,228,237,239,243,249,250,252,254,257,262,264,265,269,270,274,276,277,295,305,307,310,319,329,330,335,336,337,343,358,359,367,371,377,386,393,400,403,408,409,413,415,424,432,434,435,449,450,458,461,463,471,473,481,484,489 } Erro= 0.0000

$r\_6: [ (X51 = 1) ] \Rightarrow [ (Decisão = 1) ]$

CoverP={0,7,8,9,11,12,14,15,18,23,27,33,39,47,54,58,63,78,82,87,89,96,115,119,125,128,133,134,157,164,171,172,173,175,188,198,208,213,219,221,237,238,239,241,243,249,252,253,255,257,258,274,277,281,285,289,296,307,308,313,321,325,327,329,342,355,377,384,386,393,396,411,425,450,453,463,481,483,485,492,497 }  
Erro= 0.0123

r\_7: [ (X20 = 3) ] => [ (Decisão = 1) ]

CoverP={6,14,16,17,21,24,25,29,30,38,42,43,45,51,55,60,61,63,65,75,78,79,84,86,88,89,100,102,103,109,113,115,118,121,123,129,130,132,133,142,145,146,150,153,155,156,157,164,169,171,182,186,193,197,200,201,202,205,209,213,224,225,226,227,229,232,233,239,243,248,249,252,261,265,271,273,274,275,277,278,280,281,282,289,290,292,294,305,309,312,313,316,318,326,342,343,347,349,350,351,357,359,361,363,374,376,377,379,383,395,396,397,398,402,406,407,411,422,426,430,431,433,442,443,445,446,453,456,457,460,462,467,469,474,475,486,487,488,492,493,511,526,534 }  
Erro= 0.0210

r\_8: [ (X47 = 2) ] => [ (Decisão = 1) ]

CoverP={0,3,4,14,19,24,31,36,46,67,73,78,79,90,94,95,98,100,105,108,113,117,122,125,126,130,133,134,136,137,138,139,146,152,159,161,172,173,175,177,182,184,186,194,201,204,208,210,211,214,215,220,222,227,236,242,244,246,259,260,263,283,284,287,289,290,294,297,299,301,304,306,311,312,313,316,322,323,328,332,333,334,339,340,344,347,348,353,355,360,361,362,364,376,389,392,396,402,405,406,414,417,428,431,437,438,439,442,448,453,455,459,465,474,478,486,487,490,492 }  
Erro= 0.0000

r\_9: [ (X20 = 0), (X47 = 0), (X51 = 0) ] => [ (Decisão = 0) ]

CoverP={495,496,504,505,509,510,521,525,529,530,535,540,542,543 }  
Erro= 0.0000

r\_10: [ (X47 = 4), (X51 = 0) ] => [ (Decisão = 0) ]

CoverP={498,499,500,503,508,511,512,514,516,518,519,522,523,527,532,536,538,539,544,545 }  
Erro= 0.0000

r\_11: [ (X20 = 1), (X47 = 0), (X51 = 0) ] => [ (Decisão = 0) ]

CoverP={ 501,502,506,507,513,517,528,531,533,537 }  
Erro= 0.0000

r\_12: [ (X20 = 4), (X51 = 0) ] => [ (Decisão = 0) ]

CoverP={ 515,524 }  
Erro= 0.0000

r\_13: [ (X20 = 2), (X51 = 0) ] => [ (Decisão = 0) ]

CoverP={ 520,522,541 }  
Erro= 0.0000