

4 Resultados

Neste capítulo serão apresentados os resultados referentes à aplicação da metodologia ao banco de dados descritos nos capítulos anteriores. Na seção 4.1 serão apresentados os resultados da Seleção de Atributos e na seção 4.2, os resultados da predição do desfecho (grau de investimento ou especulação) e na seção 4.3 é apresentada uma análise comparativa entre todos os métodos de seleção e de predição utilizados.

4.1 Seleção de atributos

Primeiramente vamos analisar os resultados deste o ponto de vista da seleção de atributos para posteriormente, na seção 4.2, prosseguir na análise da predição do grau das empresas. Na tabela 5 apresenta-se os resultados comparativos entre os métodos Correlação, PCA, Informação Mútua e MIFS-U, ordenados pela relevância do atributo em relação ao desfecho (grau de investimento ou especulação) de maneira decrescente. Através desta se pode inicialmente comparar os resultados de ordenação entre correlação e IM. Pode-se observar que o conjunto dos cinco atributos mais relevantes identificados por estas duas técnicas, a saber: Lucro Líquido (9), Valor de Mercado (18), Lucro Líquido (conforme) Publicado (8), EBIT (4) e Receitas (3) são os mesmos, apesar de uma troca de ordem. Embora estes cinco atributos apresentem baixa correlação com a saída, (vide tabela 6), o valor da IM entre estes e o desfecho é alta. Este fato é um indicativo de que estes atributos estão relacionados de maneira não-linear com o desfecho, ou seja, são importantes e efetivamente relevantes na avaliação do grau de empresas, mas de um modo não-linear. Vale ressaltar que os valores numéricos da Informação Mútua e da Correlação não são passíveis de comparação direta, devido à natureza de cada método.

A partir de uma análise criteriosa sobre os resultados da tabela 6, podemos inferir que o fato dos atributos Setor (1) e Liquidez Corrente (17) terem sido

priorizados pelo modelo de seleção MIFS-U, não obstante suas baixas relevâncias com o desfecho de acordo com um critério puro de Informação Mútua (vide tabela 5), é devido a apresentarem baixa redundância (não-linear) com todos os demais atributos, fator este ratificado observando-se as suas baixas correlações lineares (vide tabela 7). Ou seja, apesar da IM de Setor e Liquidez Corrente em relação ao desfecho ser pequena, esta é importante por não estar contida em nenhum outro atributo.

Foi testado se as correlações são estatisticamente significativas através do cálculo de *p-values* para testar a hipótese de não haver correlação. Pode-se considerar cada *p-value* como sendo a probabilidade de se obter valores de correlação de mesma magnitude às encontradas neste estudo (vide tabela 8) através de uma amostra de dados aleatória, quando a correlação verdadeira é zero. Se o *p-value* é significativamente pequeno, por exemplo, menor que 0.05, então a sua respectiva correlação é significativa. Excetuando-se para os atributos Setor (1) e Liquidez Corrente (17), de modo geral, os *p-values* entre os atributos são menores que 0.05 (vide tabela 8). Logo, as correlações entre os atributos são passíveis de análise. Porém analisando-se os *p-values* entre os atributos e o desfecho (grau da empresa), estes são em sua maioria, maiores que 0.05. Aliado à este fato, vale ressaltar que todos os valores de correlação entre os atributos e desfecho são baixos (a correlação nunca ultrapassa o valor 0.17). Logo, não podemos inferir a ordenação dos atributos através da correlação linear.

Ainda analisando a tabela 5, observamos que o atributo Receitas (3) é o único presente no conjunto dos cinco atributos mais relevantes nos métodos PCA, Correlação, IM e MIFS-U reforçando a evidência de sua relevância quanto ao desfecho. No que concerne a PCA, observa-se na tabela 9 que a primeira componente principal é responsável apenas por 39.74% da variância total dos dados antes da transformação linear. Assim, para atingirmos uma meta de 95% da variância total, é necessário analisar pelo menos as cinco primeiras componentes, não obstante as duas primeiras componentes respondam por 77.41% da variância total. Este é um indicativo que o PCA pode não ser um método de seleção de atributos adequado para os dados utilizados nesta dissertação. Utiliza-se para atingirmos a meta de 95%, um procedimento de ponderação de acordo com as importâncias relativas nas componentes, conforme descrito no capítulo de metodologia. Estes resultados ponderados percentuais do PCA estão apresentados

na tabela 10, e observando-os, ratificamos a relevância do atributo Setor (1). Também segundo o método *Stepwise*, o atributo Setor (1) é selecionado, juntamente com Valor de Mercado (18) e Disponibilidade (15) resultando na seleção de apenas estes três atributos.

Relevância dos atributos com o desfecho				
Ordenação	Método de Seleção de atributos			
	Linear		Não-Linear	
	Correlação	PCA ponderado	MIFS-U	IM
1º	9	1	8	8
2º	18	5	17	9
3º	8	2	1	4
4º	4	18	3	3
5º	3	3	7	18
6º	13	7	15	15
7º	11	11	6	7
8º	2	12	5	16
9º	10	4	2	17
10º	15	14	16	5
11º	5	10	9	11
12º	6	16	11	14
13º	12	13	14	2
14º	17	15	18	6
15º	14	6	13	13
16º	16	9	12	12
17º	7	8	10	1
18º	1	17	4	10

Tabela 5 – Resultados de ordenação dos 18 atributos de acordo com a relevância em relação ao grau da empresa.

Valores resultantes de métricas de relevância em relação ao desfecho			
Código	Nome do atributo	Linear	Não-Linear
		Correlação (valor)	Informação Mútua (valor)*
8	Lucro Líquido conforme Publicado	0.16	0.19
9	Lucro Líquido	0.17	0.17
4	EBIT	0.13	0.14
3	Receitas	0.11	0.13
18	Valor de Mercado	0.17	0.10
15	Disponibilidade	-0.03	0.09
7	Despesas Financeiras	-0.14	0.09
16	Endividamento Líquido	-0.12	0.09
17	Liquidez Corrente	-0.08	0.07
5	Depreciação	-0.05	0.07
11	Ativo Corrente	0.05	0.07
14	Endividamento Total	-0.09	0.07
2	Vendas	0.03	0.06
6	Receitas Financeiras	-0.06	0.06
13	Passivo Total	0.06	0.05
12	Passivo Corrente	-0.06	0.05
1	Setor	-0.16	0.05
10	Ativo Total	-0.02	0.05

Tabela 6 – Resultados da correlação e da IM em relação ao grau de uma empresa em ordem decrescente de Informação Mútua. * Os valores da Informação Mútua e da correlação não são passíveis de comparação direta, devido à natureza de cada método.

	Setor	Vendas	Receitas	EBIT	Depreciação	Receitas Financeiras	Despesas Financeiras	Lucro Líquido Publicado	Lucro Líquido	Ativo Total	Ativo Corrente	Passivo Corrente	Passivo Total	Endividamento Total	Disponibilidade	Endividamento Líquido	Liquidez Corrente	Valor de Mercado	Grau
Setor		-0.09	-0.06	-0.07	0.04	0.11	0.02	-0.06	-0.07	-0.02	-0.04	-0.03	-0.01	-0.01	-0.03	0.00	0.07	-0.01	-0.16
Vendas			0.82	0.80	0.73	0.41	0.57	0.68	0.70	0.68	0.73	0.69	0.62	0.54	0.52	0.51	-0.15	0.71	0.03
Receitas				0.94	0.73	0.40	0.57	0.84	0.86	0.77	0.76	0.69	0.67	0.57	0.68	0.48	-0.08	0.92	0.11
EBIT					0.68	0.39	0.52	0.95	0.96	0.72	0.72	0.65	0.63	0.53	0.64	0.45	-0.07	0.87	0.20
Depreciação						0.55	0.75	0.54	0.55	0.78	0.74	0.74	0.73	0.68	0.56	0.69	-0.10	0.57	-0.05
Receitas Financeiras							0.34	0.34	0.34	0.38	0.52	0.39	0.35	0.30	0.27	0.30	0.03	0.39	0.06
Despesas Financeiras								0.32	0.32	0.93	0.86	0.94	0.96	0.98	0.83	0.98	-0.12	0.42	-0.14
Lucro Público	lo								0.99	0.54	0.56	0.47	0.44	0.37	0.48	0.26	-0.03	0.80	0.16
Lucro	lo									0.55	0.57	0.48	0.45	0.36	0.50	0.25	-0.03	0.83	0.17
Ativo											0.93	0.96	0.98	0.95	0.92	0.90	-0.11	0.68	0.02
Ativo	te											0.95	0.92	0.89	0.89	0.83	-0.03	0.67	-0.05
Passivo													0.98	0.97	0.90	0.94	-0.10	0.57	-0.06
Endividamento Total	al													0.99	0.93	0.95	-0.11	0.56	-0.06
Disponibilidade	to														0.90	0.98	-0.09	0.50	-0.09
Endividamento Líquido																0.80	0.02	0.65	-0.03
Liquidez Corrente																	-0.14	0.33	-0.12
Valor de Mercado																		-0.03	-0.08
Grau																			0.17

Tabela 7 – Matriz de correlações entre todos os dezoito atributos e o grau de uma empresa.

		Setor	Vendas	Receitas	EBIT	Depreciação	Receitas Financeiras	Despesas Financeiras	Lucro Líquido Publicado	Lucro Líquido	Ativo Total	Ativo Corrente	Passivo Corrente	Passivo Total	Endividamento Total	Disponibilidade	Endividamento Líquido	Liquidez Corrente	Valor de Mercado	Grau
Setor			0.12	0.31	0.21	0.43	0.04	0.75	0.25	0.22	0.75	0.47	0.58	0.83	0.88	0.63	0.00	0.20	0.88	0.00
Vendas				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.65
Receitas					0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.00	0.06
EBIT						0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24	.00	0.02
Depreciação							0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.35
Receitas Financeiras								0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.54	0.00	0.30
Despesas Financeiras									0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.01
Lucro Líquido Publicado	lo									0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.56	0.00	0.00
Lucro Líquido	lo										0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.55	0.00	0.00
Ativo Total												0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.73
Ativo Corrente	te												0.00	0.00	0.00	0.00	0.00	0.60	0.00	0.42
Passivo Corrente														0.00	0.00	0.00	0.00	0.08	0.00	0.29
Passivo Total	al														0.00	0.00	0.00	0.05	0.00	0.27
Endividamento Total	to															0.00	0.00	0.10	0.00	0.10
Disponibilidade																	0.00	0.69	0.00	0.64
Endividamento Líquido																		0.01	0.00	0.04
Liquidez Corrente																			0.65	0.16
Valor de Mercado																				0.00
Grau																				

Tabela 8 – Matriz de *p-values* entre todos os dezoito atributos e o grau de uma empresa.

Importância das componentes principais		
Componente	Importância da componente (%)	% acumulado
1 ^a	39.74	39.74
2 ^a	37.67	77.41
3 ^a	8.17	85.58
4 ^a	6.44	92.02
5 ^a	3.18	95.20
6 ^a	1.87	97.07
7 ^a	1.32	98.39
8 ^a	0.89	99.28
9 ^a	0.23	99.51
10 ^a	0.23	99.74
11 ^a	0.09	99.83
12 ^a	0.07	99.90
13 ^a	0.04	99.94
14 ^a	0.03	99.97
15 ^a	0.02	99.99
16 ^a	0.01	100.00
17 ^a	0.00	100.00
18 ^a	0.00	100.00

Tabela 9 – Importância percentual e acumulada das componentes principais.

Relevâncias dos atributos considerando as componentes ponderadas		
Ordenação	Código do atributo	Relevância ponderada (%)
1º	1	16.46
2º	5	6.67
3º	2	6.64
4º	18	6.31
5º	3	6.22
6º	7	5.33
7º	11	5.32
8º	12	5.24
9º	4	4.84
10º	14	4.75
11º	10	4.65
12º	16	4.65
13º	13	4.65
14º	15	3.93
15º	6	3.77
16º	9	3.75
17º	8	3.61
18º	17	3.19

Tabela 10 – Relevância dos atributos em relação ao desfecho ponderados pela importância percentual das componentes principais.

4.2

Predição do grau da empresa

Nesta seção, serão analisados os experimentos referentes à predição do desfecho (grau de investimento ou especulação) através dos métodos de Redes Neurais, Regressão Múltipla Linear e Discriminante Linear de Fisher. Os experimentos realizados têm como objetivo comparar a capacidade de predição do desfecho destes métodos auxiliados pelos métodos de seleção de atributos descritos na seção 3.1 e por conhecimento a priori.

Os resultados nos remeteram a três questões conclusões interessantes:

1. Alguns métodos de seleção de atributos, quando aplicados, resultam em uma melhor eficácia na posterior predição do desfecho. Contudo, as diferenças entre os resultados produzidos pelos diversos métodos não se mostrou significativa;
2. No modelo de predição através de Redes Neurais o porte e o setor da empresa se mostraram bastante importantes;
3. Métodos lineares não são bons preditores para o desfecho, reforçando a idéia de que este problema necessita de um modelo não-linear.

4.2.1

Predição através de Redes Neurais

4.2.1.1

Predição após selecionar os atributos de acordo com a Informação Mútua (IM).

O conjunto dos atributos mais relevantes em relação ao desfecho selecionados através do método de Informação Mútua e composto por Lucro Líquido Publicado (8), Lucro Líquido (9), EBIT (4), Receitas (3), Valor de Mercado (18), Disponibilidade (15) e Despesas Financeiras (7), apresentou o melhor resultado para a predição do grau das empresas com uma taxa de acerto médio de 91.86% (vide tabela 11 e figura 2). Este resultado supera o resultado obtido com a utilização de todos os 18 atributos demonstrando a alta relevância daqueles sete atributos principais em relação ao desfecho e a eficácia da utilização da IM como método de seleção de atributos para a predição do desfecho. Pode-se observar que conforme o conjunto de atributos aumenta (os atributos são inseridos de acordo com a ordenação pela IM), a taxa de acerto para o grau de especulação aumenta mais rapidamente do que a taxa de acerto para o grau de investimento (vide tabela 11). É interessante ressaltar que o conjunto composto pelos atributos Lucro Líquido Publicado (8), Lucro Líquido (9), EBIT (4), Receitas (3) e Valor de Mercado (18), proporcionou um aumento de mais de 13% na taxa de predição do desfecho em relação ao conjunto anterior (vide tabela 10) composto pelos atributos Lucro Líquido Publicado (8), Lucro Líquido (9), EBIT (4), Receitas (3).

Este é um indicativo de que o tamanho da empresa, avaliado pelo seu valor de mercado, é muito relevante para a atribuição do grau da empresa pelas agências classificadoras de risco.

Ordenação IM				
Atributos	Resultados (2000 épocas e 10 neurônios iniciais)			
	Acerto (%) Especulação	Acerto (%) Investimento	Acerto (%) Médio	Número de parâmetros efetivos
8	67.86	71.37	69.62	13.32
8, 9	71.43	75.19	73.31	11.73
8, 9, 4	73.21	79.39	76.30	16.47
8, 9, 4, 3	75.00	79.01	77.01	41.34
8, 9, 4, 3, 18	94.64	87.02	90.83	58.62
8, 9, 4, 3, 18, 15	91.07	85.11	88.09	40.88
8, 9, 4, 3, 18, 15, 7	98.21	85.50	91.86	48.27
8, 9, 4, 3, 18, 15, 7, 16	94.64	88.55	91.60	56.77
8, 9, 4, 3, 18, 15, 7, 16, 17	94.64	86.64	90.64	47.21
Todos os dezoito atributos	94.29	88.17	91.23	57.6490

Tabela 11 – Resultados de predição do grau de empresas utilizando os atributos ordenados através do método de IM.

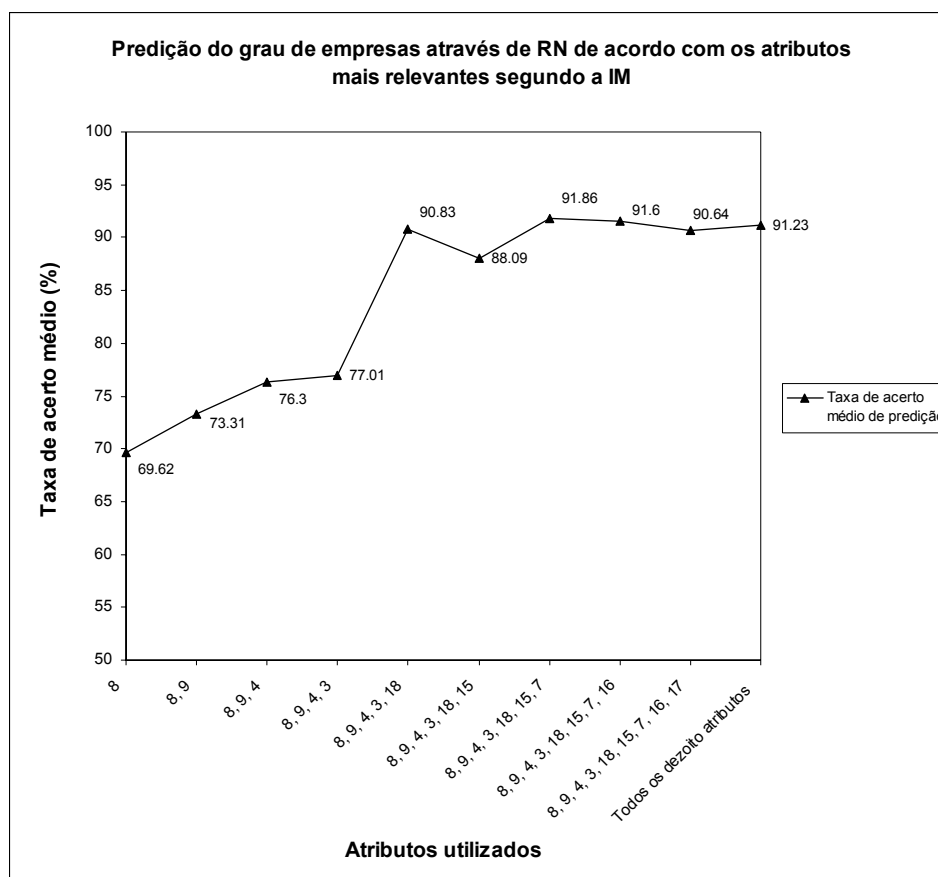


Figura 2 – Comportamento da evolução da taxa de acerto médio através dos atributos inseridos baseados na ordenação resultante do método de IM.

4.2.1.2

Predição após selecionar os atributos de acordo com o MIFS-U.

O Conjunto de atributos mais relevantes em relação ao desfecho selecionadas através do método MIFS-U e composto por Lucro Líquido Publicado (8), Liquidez Corrente (17), Setor (1), Receitas(3) e Despesas Financeiras(7), apresentou o melhor resultado para a predição do grau das empresas fazendo o uso posterior de Redes Neurais com taxa de acerto média de 91.14% (vide tabela 12 e figura 3). Apesar de marginalmente inferior ao resultado obtido com a utilização de todos os atributos, há uma clara indicação de que o conjunto de atributos (8, 17, 1, 3, 7) selecionados dada à ordenação de relevância quanto ao desfecho pelo MIFS-U é importante para a definição do grau de empresas.

Também pode-se observar que conforme o conjunto de atributos aumenta (os atributos são inseridos de acordo com a ordenação pelo MIFS-U), a taxa de acerto para o grau de especulação aumenta mais rapidamente do que a taxa de

acerto para o grau de investimento (vide tabela 12), tal como o que ocorre quando os atributos são inseridos de acordo com a ordenação pela IM (vide tabela 11). Este comportamento da evolução distinta das taxas de acerto para o grau de especulação frente as do grau de investimento, pode indicar que é necessário a análise de um número maior de atributos para classificar as corporações como grau de especulação.

Ordenação MIFS-U				
Atributos	Resultados (2000 épocas e 10 neurônios iniciais)			
	Acerto (%) Especulação	Acerto (%) Investimento	Acerto (%) Médio	Número de parâmetros efetivos
8	67.86	71.37	69.62	13.3214
8, 17	87.50	74.43	80.97	47.0626
8, 17, 1	96.43	75.57	86.00	68.3084
8, 17, 1, 3	94.64	77.10	85.87	54.3551
8, 17, 1, 3, 7	96.78	84.88	91.14	52.0028
8, 17, 1, 3, 7, 15	96.43	82.82	89.63	52.9799
8, 17, 1, 3, 7, 15, 6	92.86	83.59	88.23	59.8400
Todos os dezoito atributos	94.29	88.17	91.23	57.6490

Tabela 12 – Resultados de predição do grau de empresas utilizando os atributos ordenados através do método de MIFS-U.

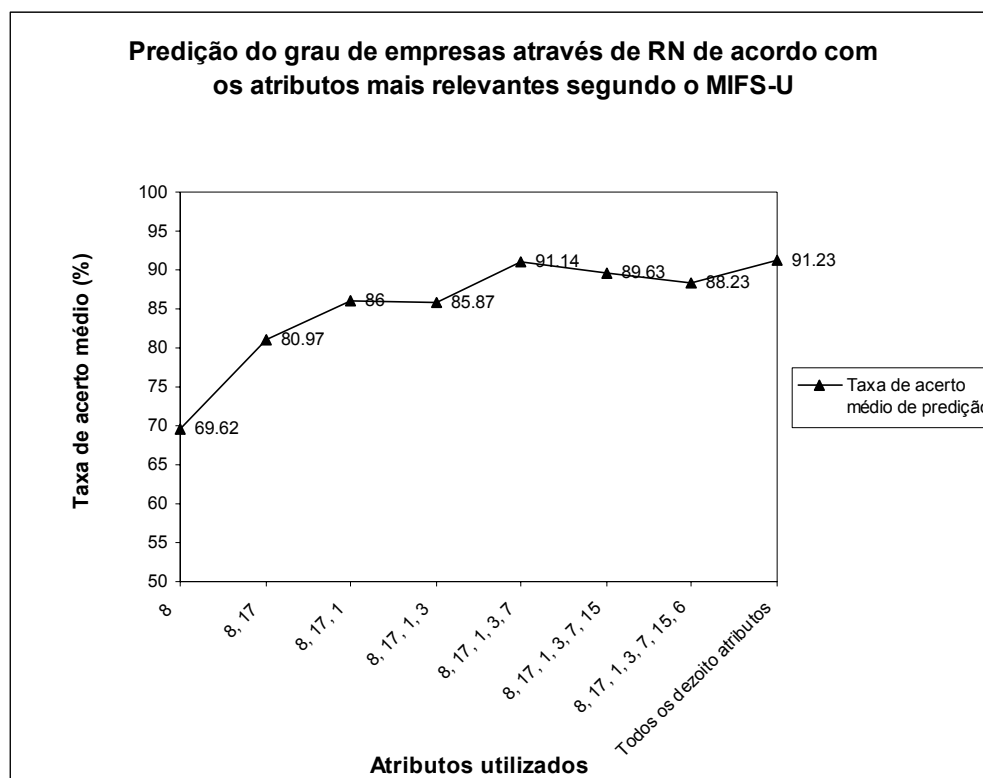


Figura 3 – Comportamento da evolução da taxa de acerto médio através dos atributos inseridos baseados na ordenação resultante do método de MIFS-U.

Visando comprovarmos a robustez do resultado obtido através de RNs para os atributos selecionados pelo MIFS-U *Lucro Líquido (conforme) Publicado* (8), *Liquidez Corrente* (17), *Setor* (1), *Receitas* (3) e *Despesas Financeiras*(7), executamos o procedimento k-fold, ($k=5$). O k-fold executa k simulações da mesma arquitetura da Rede Neural, incluindo suas respectivas validações e treinamentos, porém inicializados independentemente de maneira que os pesos aleatórios iniciais das k Redes Neurais são distintos em cada uma das k simulações (vide tabela 13 e figura 4). Este conjunto de atributos foi escolhido por ser o que apresentou o melhor resultado para a predição do desfecho utilizando o método de seleção MIFS-U, demonstrando-se robusto, dado a média do resultado de predição do desfecho para ambos os graus e seu baixo desvio-padrão.

Resultados 5-fold com os atributos 8, 17, 1, 3, 7				
Simulação	(2000 épocas e 10 neurônios iniciais)			
	Acerto (%) Especulação	Acerto (%) Investimento	Acerto (%) Médio	Número de parâmetros efetivos
1	96.43	85.11	92.27	51.27
2	98.21	83.59	90.90	55.87
3	92.86	85.88	89.37	45.52
4	98.21	83.59	90.90	56.92
5	98.21	86.26	92.24	50.41
Média	96.78	84.88	91.14	52.00
Desvio-padrão	2.33	1.25	1.19	4.59

Tabela 13 – Resultados provenientes de 5 simulações da RN para a predição do grau de empresas utilizando os atributos 8, 17, 1, 3 e 7 selecionados baseado na ordenação através do método de MIFS-U.

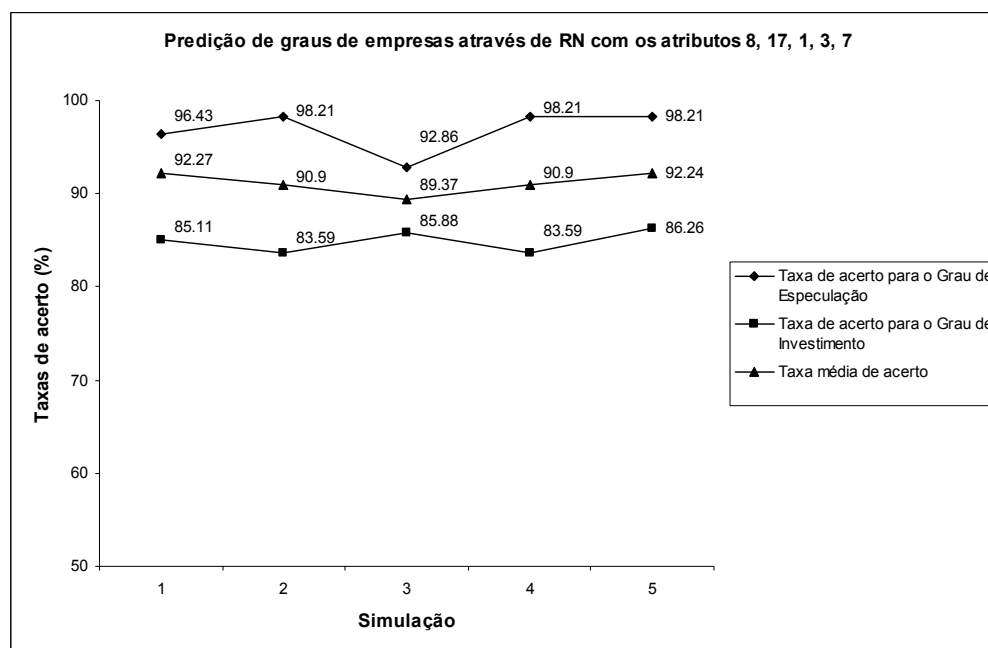


Figura 4 – Comportamento da evolução da taxa de acerto em 5 simulações para os graus de especulação e investimento utilizando aos atributos 8, 17, 1, 3 e 7 selecionados baseado na ordenação resultante do método de MIFS-U.

4.2.1.3

Predição após selecionar os atributos de acordo com o PCA

O Conjunto de atributos mais relevantes em relação ao desfecho selecionados através do método PCA ponderada a importância dos atributos nas componentes (descrito na seção de metodologia), e composto por Setor (1), Vendas (2), Receitas (3), Depreciação (5) e Valor de Mercado (18) e Despesas Financeiras (7) apresentou o melhor resultado para a predição do grau das empresas fazendo o uso posterior de Redes Neurais com taxa de acerto média de 90.01%. Este resultado, apesar de próximo, é abaixo do resultado obtido com a utilização de todos os atributos, cuja taxa de acerto média é de 91.23%. (vide tabela 14 e figura 5). Portanto, caberia na utilização destes métodos, uma análise da diminuição da necessidade de acesso a dados e do menor tempo de processamento da Rede Neural, frente a menor taxa de acerto da obtida com a utilização dos dezoito atributos.

Ordenação PCA ponderado

Atributos	Resultados (2000 épocas e 10 neurônios iniciais)			
	Acerto (%) Especulação	Acerto (%) Investimento	Acerto (%) Médio	Número de parâmetros efetivos
1	75.00	50.00	62.50	8.28208
1, 5	42.86	72.90	57.88	12.7296
1, 5, 2	73.21	67.18	70.20	43.9520
1, 5, 2, 18	96.43	80.15	88.29	52.0139
1, 5, 2, 18, 3	98.21	81.30	89.76	62.6581
1, 5, 2, 18, 3, 7	96.43	83.59	90.01	51.3392
1, 5, 2, 18, 3, 7, 11	91.07	88.55	89.81	45.388
Todos os dezoito atributos	94.29	88.17	91.23	57.6490

Tabela 14 – Resultados de predição do grau de empresas utilizando os atributos ordenados através do método PCA ponderado.

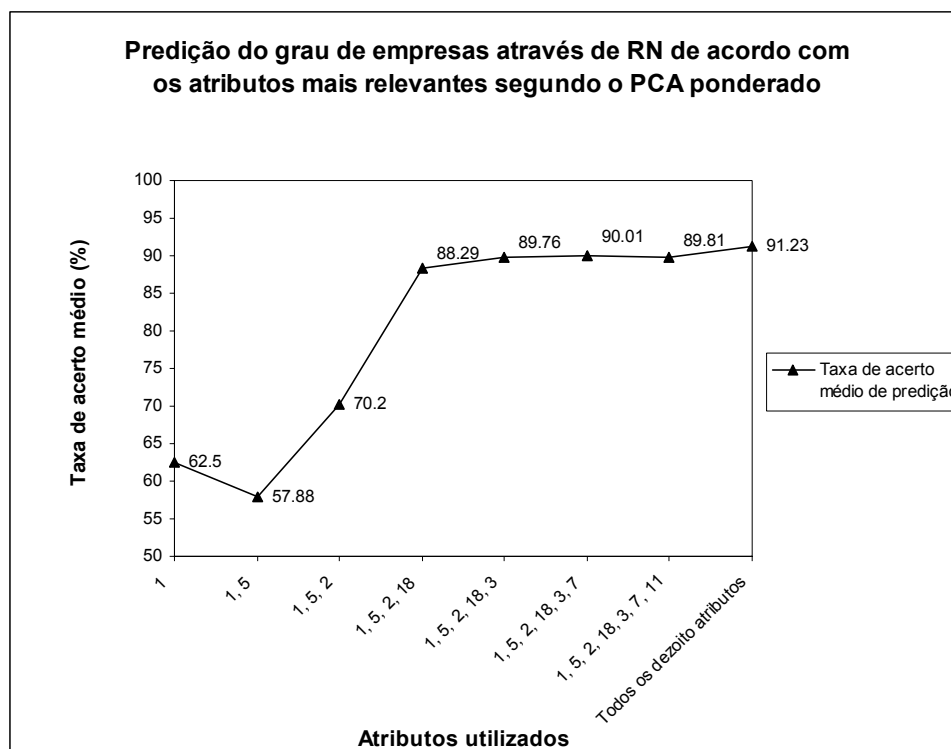


Figura 5 – Comportamento da evolução da taxa de acerto médio através dos atributos inseridos baseados na ordenação resultante do método de PCA ponderado.

4.2.1.4

Predição após selecionar os atributos com o *Stepwise*

O Conjunto de atributos mais relevantes em relação ao desfecho selecionados através do método *Stepwise* é o composto pelos atributos Valor de Mercado (18), Disponibilidade (15) e Setor (1). Utilizando este conjunto e fazendo o uso posterior de Redes Neurais obteve-se uma taxa de acerto média de 88.16% para a predição do desfecho (vide experimento No. 1 na tabela 18). Este resultado e o resultado do PCA ponderado foram os dois resultados mais baixos obtidos entre todos os métodos de seleção de atributos com a utilização posterior de Redes Neurais para a predição do desfecho. Este fato pode indicar que os métodos lineares de seleção de atributos não são os mais adequados, dentre os estudados, para a predição de graus de empresas. Observa-se, portanto, que as relações entre os atributos e o desfecho parecem ser não-lineares.

4.2.1.5

Predição após selecionar os atributos de acordo com a Informação Mútua, correlação e conhecimento a priori

Analisando a tabela 5 observamos que os cinco atributos detentores das mais altas relevâncias com o desfecho segundo a Informação Mútua são: Lucro Líquido Publicado (8), Lucro Líquido (9), EBIT (4), Receitas(3), e Valor de Mercado (18). Estes atributos são muito importantes para se avaliar o porte (tamanho) da empresa, e identificar o quão grande é a empresa e considerar seu setor de atuação é fundamental para se determinar seu grau, uma vez que grandes corporações e/ou de setores industriais em ascensão enfrentam um risco menor de falência do que empresas pequenas e/ou atuantes em setores em decadência (TREACY et. al 2000). Ou seja, a área de atuação da empresa pode de fato potencializar a exposição à impactos econômicos setoriais devido à mudanças competitivas estruturais em andamento naquele tipo de indústria (CANTOR 2005). De posse do conhecimento a priori do atributo Setor (1) e de seu destaque como relevante obtida principalmente nos métodos de seleção de atributos PCA e MIFS-U, decidiu-se utilizá-lo com os atributos que expressam em sua natureza o porte da empresa e de mais alta relevância segundo a Informação Mútua. Também se observou cujos dois atributos mais relevantes segundo este método (Lucro Líquido (9) e Lucro Líquido (conforme) Publicado (8)) apresentam alta correlação linear (0.99) entre si (vide tabela 7). Portanto, realizou-se o experimento de No. 5 com os atributos Lucro Líquido (9), EBIT (4), Receitas (3), Valor de Mercado (18) e Setor (1) e o experimento de No. 6 com os atributos Lucro Líquido Publicado (8), EBIT (4), Receitas(3), Valor de Mercado (18) e Setor (1). Ambos experimentos são apresentados na tabela 18.

O conjunto constituído pelos cinco atributos do experimento No. 5, a saber: Lucro Líquido (9), EBIT (4), Receitas (3), Valor de Mercado (18) e Setor (1) foi capaz de estimar o grau das empresas com um resultado de 100% de acerto médio na predição específica do grau de especulação das empresas, resultado este excelente e robusto dado 5 simulações do k-fold (vide tabela 15 e figura 6). Para o grau de investimento o resultado médio de acerto do grau de empresas nas mesmas 5 simulações do k-fold foi de 88.63%. A existência de relações não-lineares (relevância) entre os atributos do experimento No. 5 e o desfecho é então

ratificada pelo excelente índice de acerto médio na predição do desfecho através do modelo de Redes Neurais capaz de capturar estas relações.

O conjunto de atributos do experimento No. 6, a saber: Lucro Líquido (conforme) Publicado (8), EBIT (4), Receitas (3), Valor de Mercado (18) e Setor (1) obteve um resultado médio de acerto na predição do desfecho de 91.66%, sendo este o terceiro melhor resultado de todos os experimentos realizados. A diferença entre os resultados obtidos no experimento No. 6 e de No. 5 é um indicativo de que as agências de risco consideram mais relevante o lucro total da empresa, do que o lucro líquido segundo normas contábeis norte-americanas para a atribuição de seu grau.

Resultados k-fold com os atributos 1, 9, 4, 3, 18				
Simulação	(2000 épocas e 10 neurônios iniciais)			
	Acerto (%) Especulação	Acerto (%) Investimento	Acerto (%) Médio	Número de parâmetros efetivos
1	100	88.17	94.09	50.2876
2	100	87.40	93.70	55.0897
3	100	90.08	95.04	56.5779
4	100	89.31	94.66	48.7341
5	100	88.17	94.09	50.1919
Média	100	88.63	94.32	52.1762
Desvio-padrão	0	1.06	0.53	3.4357

Tabela 15 - Resultados provenientes de 5 simulações da RN para a predição do grau de empresas utilizando os atributos 1, 9, 4, 3, 18.

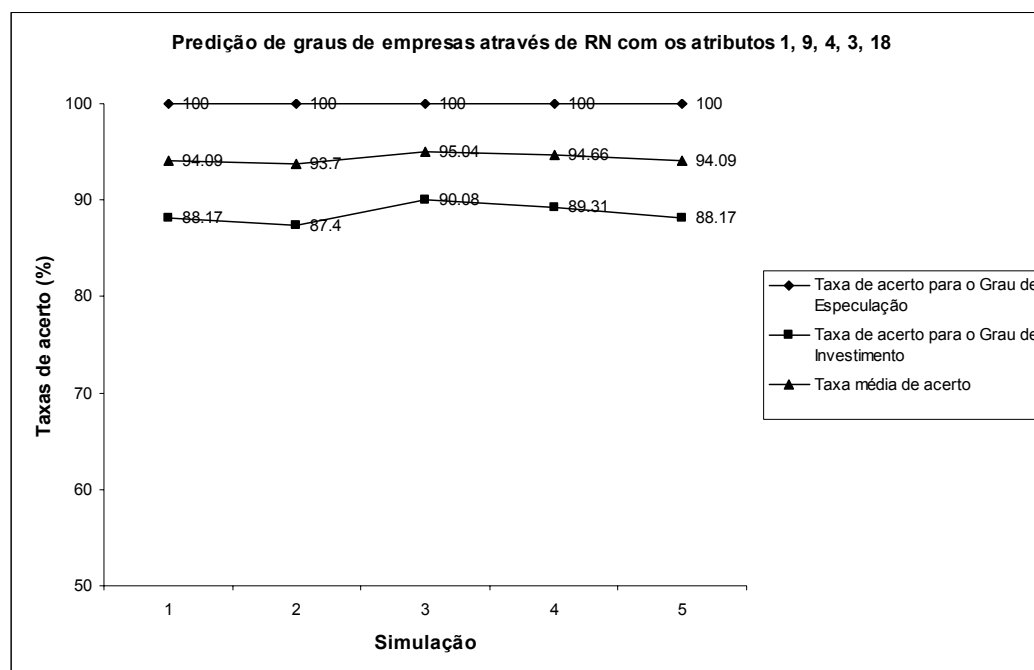


Figura 6 - Comportamento da evolução da taxa de acerto para os graus de especulação e investimento utilizando os atributos 1, 9, 4, 3, 18.

4.2.2

Predição através de Métodos Lineares

Os resultados de acerto médio do desfecho através de métodos lineares foram significativamente menores comparando-os com a predição através de Redes Neurais e seus valores máximos foram obtidos fazendo o uso de todos os atributos: O método de Regressão Múltipla Linear obteve uma taxa máxima de 76% de acerto médio e o método de Discriminante Linear de Fisher obteve uma taxa máxima de 72.4% de acerto médio (vide tabela 18 e figura 7). Pode-se observar os coeficientes do modelo de regressão linear utilizado e um sumário de testes estatísticos do modelo de regressão utilizado nas tabelas 16 e 17 respectivamente. Os resultados desta seção reforçam a observação de que as relações entre os atributos e o desfecho (grau das empresas) são essencialmente, não-lineares.

Sumário do Modelo de Regressão Linear			
R	R^2	R^2 Ajustado	Desvio-Padrão
0.58	0.26	0.23	0.44

Tabela 16 – Testes estatísticos da regressão linear

Atributo	Coeficientes		Coeficientes Padronizados	t	Intervalo de Confiança de 95% para B	
	B	Desvio-Padrão	β		Limite inferior	Limite superior
Setor	-0.11	0.03	-0.14	-3.32	-0.18	-0.05
Vendas	-0.52	0.28	-0.23	-1.85	-1.06	0.03
Receitas	0.13	0.49	0.05	0.26	-0.83	1.08
EBIT	0.39	0.92	0.11	0.43	-1.41	2.19
Depreciação	-0.32	0.24	-0.18	-1.29	-0.80	0.16
Receitas Financeiras	0.04	0.20	0.02	0.22	-0.36	0.44
Despesas Financeiras	-4.05	0.81	-1.93	-4.99	-5.65	-2.46
Lucro Líquido Publicado	-0.03	1.55	-0.01	-0.02	-3.07	3.01
Lucro Líquido	-0.17	1.87	-0.04	-0.09	-3.85	3.51
Ativo Total	4.01	2.18	1.33	1.84	-0.27	8.29
Ativo Corrente	0.19	0.62	0.08	0.31	-1.03	1.41
Passivo Corrente	-0.37	1.19	-0.16	-0.31	-2.71	1.96
Passivo Total	-3.99	3.02	-1.42	-1.32	-9.92	1.95
Endividamento Total	-4.06	2.98	-1.49	-1.36	-10.03	2.54
Disponibilidade	-1.84	1.16	-0.45	-1.58	-4.12	0.44
Endividamento Líquido	5.78	1.46	2.52	3.95	2.90	8.66
Liquidez Corrente	-0.09	0.08	-0.05	-1.17	-0.24	0.06
Valor de Mercado	1.50	0.39	0.50	3.81	0.73	2.27

Tabela 17 – Coeficientes da Regressão Linear

4.3

Análise comparativa dos resultados obtidos através dos métodos de seleção e predição

Observando o resultado dos experimentos realizados com o PCA e o *Stepwise* com a utilização posterior de Redes Neurais para a predição do desfecho respectivamente, experimento No. 1 e de No. 2 (vide tabela 18 e figura 7) verifica-se que estes são os resultados de predição mais baixos dentre os experimentos realizados com Redes Neurais. Estes resultados são indicativos de que métodos de seleção de atributos lineares são menos eficazes para a predição do desfecho (grau de especulação ou investimento) principalmente quando utilizados antes de métodos de predição não-lineares.

Analisando o resultado de predição de 91.14% do experimento de No. 3 (vide tabela 18 e figura 7) observa-se que apenas supera os resultados obtidos através dos métodos de seleção lineares, sendo o quinto melhor resultado dos experimentos comparativos realizados apresentados.

Comparando o resultado de predição do experimento de No. 4 com os demais obtidos através de Redes Neurais, observa-se que este é o segundo melhor resultado de todos os experimentos realizados, somente atrás do experimento de No. 5. Uma vez que o experimento No. 5 faz uso de uma análise que utiliza o conhecimento *a priori* da importância do atributo Setor (1), este é o método que apresentou o melhor resultado exclusivamente através da aplicação de métodos de seleção de atributos e de predição.

A diferença entre os resultados obtidos no experimento No. 5 e de No. 6 é um indicativo de que as agências de risco consideram mais relevante o lucro total da empresa, do que a Lucro Líquido segundo normas contábeis norte-americanas (Lucro Líquido Publicado) para a atribuição de seu grau.

Observando exclusivamente os resultados de predição obtidos através de métodos lineares, observa-se que a utilização dos métodos de seleção de atributos estudados não resultaram em um grande ganho para a predição do grau de empresas uma vez que a grande maioria dos resultados utilizando estes métodos foram significativamente menores do que os obtidos com a utilização de todos os atributos. A única exceção está no experimento No. 4 (vide tabela 18), cujo método de seleção utilizado foi o baseado na Informação Mútua seguido pelo método de predição Discriminante Linear de Fisher, resultando em uma taxa de acerto média de predição de 72.3% (próxima à taxa de 72.4% com a utilização de todos os atributos).

Porém, de forma geral, os resultados de predição obtidos através de Redes Neurais com a utilização de sub-conjuntos (Experimentos Nos. 1, 2, 3, 4, 5 e 6) identificados através dos métodos de seleção de atributos se aproximam dos resultados obtidos com a utilização de todos os atributos (Experimento No. 7). Desta maneira, de posse destes sub-conjuntos (vide tabela 18), pode-se obter um bom índice de acerto médio do desfecho, impactando na necessidade de obtenção e análise de volumes de dados menores e da maior eficiência da respectiva Rede Neural. Também com base nos experimentos realizados pode-se ratificar a natureza não-linear das relações entre os atributos e do grau da empresa;

Método de Seleção de Atributos				Método de Predição		
No. Exp.	Modelo	Natureza	Atributos utilizados	Linear		Não-linear
				Regressão Múltipla	Discriminante de Fisher	Redes Neurais MLP
				Acerto médio (%)	Acerto médio (%)	Acerto médio (%)
1	PCA ponderado	Linear	1, 2, 3, 5, 18, 7	67.90	68.60	90.01
2	Stepwise	Linear	18, 15, 1	67.70	68.60	88.16
3	MIFS-U	Não-linear	8, 17, 1, 3, 7	66.20	65.50	91.14
4	IM	Não-linear	8, 9, 4, 3, 18, 15, 7	68.10	72.30	91.86
5	IM + correlação + informação a priori	Não-linear	1, 9, 4, 3, 18	67.70	65.10	94.32
6	IM + correlação + informação a priori	Não-linear	1, 8, 4, 3, 18	69.00	66.10	91.66
7	Nenhum	-	Todos os atributos	76.00	72.40	91.23

Tabela 18 – Matriz comparativa entre métodos de seleção lineares e não lineares com a utilização posterior de métodos de classificação lineares e não-lineares.

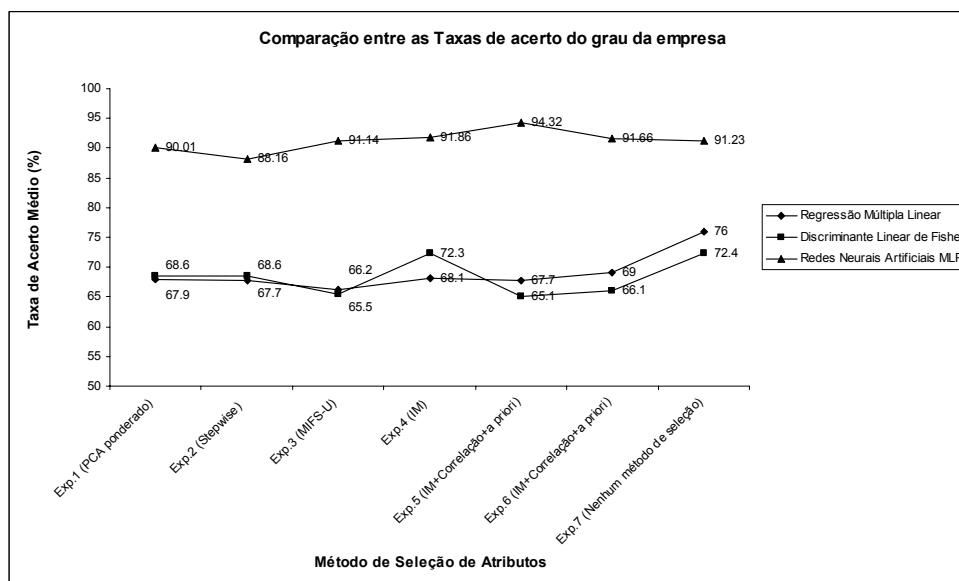


Figura 7 – Comportamento da variação das taxas de acerto médio com a utilização dos métodos de seleção de atributos e predição de desfecho (grau de investimento ou especulação).