UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ADMINISTRAÇAO
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO

Humberto Carlos L'Astorina

USANDO REDES BAYESIANAS PARA A PREVISÃO DA RENTABILIDADE DE EMPRESAS

Porto Alegre
2009

Humberto Carlos L'Astorina

USANDO REDES BAYESIANAS PARA A PREVISÃO DA RENTABILIDADE DE EMPRESAS

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Administração da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do título de Mestre em Administração.

Orientador: Prof. Dr. Denis Borenstein

Porto Alegre
2009

Dados Internacionais de Catalogação na Publicação (CIP)

L349u L'Astorina, Humberto Carlos

Usando redes Bayesianas para a previsão da rentabilidade de empresas / Humberto Carlos L'Astorina. – 2009.

92 f. : il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul, Escola de Administração, Programa de Pós-Graduação em Administração, 2009.

Orientador: Prof. Dr. Denis Borenstein.

1. Previsões. 2. Rentabilidade futura. 3. Redes Bayesianas. I. Título.

CDU 681.3

Ficha elaborada pela equipe da Biblioteca da Escola de Administração UFRGS

SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ADMINISTRAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO

BANCA EXAMINADORA:

Prof. Dr. João Luiz Becker

(PPGA/EA/UFRGS)

Prof. Dr. Eduardo Ribas Santos

(EA/UFRGS)

Prof. Dr. João Carlos Furtado

(UNISC)

Orientador: Prof. Dr. Denis Borenstein

Área de Concentração: Sistemas de Informação e de Apoio à Decisão

Curso: Mestrado Acadêmico

Porto Alegre, 27 de Abril de 2009.

Dedico este trabalho a meus pais Edith e Ednyr, que colocaram a educação dos filhos sempre em primeiro lugar. À Carla, amor de minha vida, que me apóia , ajuda e motiva. Às minhas filhas, fontes inesgotáveis de inspiração, Julia e Sofia. Essas, num futuro, ao lerem este trabalho provavelmente dirão com uma expressão de confirmação: “Papai sempre teve um jeito diferente de se divertir”.

AGRADECIMENTOS

Agradeço aos amigos-professores Doutores João Luiz Becker, Eduardo Ribas Santos e Denis Borenstein por me aceitarem como aluno desta escola e por terem me proporcionado a honra de compartilhar seus conhecimentos. Ao último, na função de orientador, o hífen que une as palavras “amigo” e “professor” é, sem dúvida, em negrito pela paciência e sabedoria, às vezes silenciosa, que me conduziu até o final do curso. A todos eles, meus mais sinceros votos de admiração e respeito.

Ao amigo Ary B. Calovi, que na minha ausência dos negócios cuidou desses com zelo e dedicação. Sem ele meus estudos seriam praticamente impossíveis.

Por fim, ao povo brasileiro que financia esta escola, que é do mais alto nível profissional e científico, por ter contribuído com seus impostos líquidos e certos na esperança, somente provável, de obter um trabalho de qualidade.

RESUMO

O presente trabalho emprega Redes Bayesianas para a previsão da rentabilidade de empresas. Define-se como rentabilidade superior as empresa que obtiveram retorno para os acionistas classificados acima de 81,5% em relação às demais. Adota-se a metodologia de seleção dos indicadores proposta por Sun e Shenoy (2007), que seleciona as variáveis explicativas segundo suas correlações com a variável classificadora. Obtêm-se, ao final, dois modelos sendo o primeiro com dois estados de classificação de empresas, superior e inferior; o segundo com três estados (superior mediano e inferior). Assim como Sun e Shenoy (2007), tenta-se validar o modelo Bayesiano com a regressão logística. Constata-se que não é possível afirmar que as média das taxas de sucesso dos dois modelos sejam diferentes ao se prever rentabilidade superior, entretanto a regressão tem melhor desempenho ao se prever rentabilidade baixa. A variável mais significativa tanto para o primeiro quanto para o segundo modelos foi a classificação atual da empresa, ou seja, empresas que figuram em um determinado ano no estado de rentabilidade superior são as mais propensas a repetir o resultado do que as demais. Os resultados apontam taxas de acerto que vão de 14,70% em 1999 (ano da crise cambial quando a rentabilidade média das empresas foi de 2,74%) a 52,94% em 1997 (ano cuja rentabilidade média foi de 11,76%) para o primeiro modelo e de 11,76 % (1999) a 56,60 % (2004, rentabilidade média de 10,76%) para o segundo modelo. Apesar dos modelos ainda não conseguirem alcançar uma estabilidade nas previsões os resultados são animadores quando se desenvolve a hipótese de utilidade para um possível investidor e a expectativa de retorno acumulado, ao longo dos dez anos, passa de 70,37%, que é a rentabilidade média acumulada do período, para 357,07% e 410,10 % para o primeiro e o segundo modelo respectivamente.

Palavras-chave: Previsões. Rentabilidade Futura. Redes Bayesianas.

ABSTRACT

This work use the knowledge obtained from Bayesian networks studies of bankruptcy prediction and applied it for forecasting companies' profitability. Higher profitability is defined as the company that had returns for shareholders classified over 81.5% compared to the others. Adopting the methodology of selection of the explanatory variables proposed by Sun and SHENOY (2007) based on correlations among them with the classification variable. As a result it is obtained two models, the first one with two classification states for de classification variable, upper and low, and the second one with three states (upper, middle and low). As Sun and SHENOY (2007), the Bayesian model was compared with a logistic regression. It cannot be say that the average success rates of the two models are different for forecasting higher profitability; otherwise, for low profitability forecasts the regression model was superior. The most significant variable for both the first and for the second model was the previous company's return for the shareholders, i.e. companies that are in a given year in the state of upper profitability are more likely to repeat the resulting the next year. The results show success rates ranging from 14.70% in 1999 (year of the currency crisis when the average profitability of the companies was 2.74%) to 52.94% in 1997 (average return rate was 11.76 %) for the first model and from 11.76% (1999) to 56.60% (2004, average return rate was 10.76%) for the second model. Although the models still fail to achieve stability in the estimates the results are encouraging when developing the hypothesis of possible investor profitability when the expectation of return accumulated over the ten years, range from 70.37%, which is the average profitability accumulated in the period to 357.07% and 410.10% respectively for the first and second model.

Keywords: Forecast. Future Profitability. Bayesian Networks.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

Prever as rentabilidades futuras das empresas pode significar a diferença entre o sucesso e o fracasso de qualquer investidor principalmente em momentos de incerteza como o que passa a econômica mundial. A maioria dos estudos acadêmicos, publicados até o momento, abordam a probabilidade de falência das empresas, mas não a probabilidade de se obter lucros acima da média geral. O presente trabalho utiliza os conhecimentos obtidos com os estudos sobre previsão de falências e os aplica para a previsão de rentabilidades futuras das empresas adotando como variável classificadora o Retorno ao Acionista (RA).

O retorno aos acionistas (RA) médio acumulado das empresas brasileiras no período de 1998 até 2007 foi de 70,34 % enquanto o RA médio das empresas que figuraram na lista das mais rentáveis foi de 524%. Empresas como a Souza Cruz, Lojas Americanas e Weg tiveram um RA acumulado de 4.9241%, 1.527 % e 1.255% respectivamente (Tabela 1). Dispor, portanto, de uma ferramenta capaz de identificar essas empresas ou mesmo nortear futuros investimentos que ofereça um grau razoável de precisão, pode representar um incremento significativo de conhecimento e consequente vantagem competitiva de um investidor frente aos demais seja na decisão de aporte de capital ou mesmo como mecanismo de avaliação para fusões e aquisições.

Após revisão da bibliografia sobre o assunto, foi constatado que os modelos que mais se aproximaram da resposta a este problema foram os modelos de previsão de falências. Tais modelos são úteis para a análise de risco ao investidor, para estimular iniciativas das autoridades monetárias no sentido de intervenção em instituições financeiras e para serem empregados como uma ferramenta gerencial que indique a necessidade da correção de rumos da empresa para que esta não venha a enfrentar uma falência futura e utilizam dados das empresas correlacionados com a falência ocorridas no passado para determinar as chances de isso ocorrer no futuro .

Os modelos de previsão de falências são estudados desde a década de 1930 e são várias as técnicas empregadas. Kumar e Ravi (2006) em sua revisão sobre esses estudos apresentam 74 trabalhos publicados, somente no período de 1968 até 2005. O número significativo de trabalhos sobre o tema demonstra, de antemão, sua importância.

Tabela 1 - Rentabilidade ao acionista acumulada das empresas mais rentáveis no período de 1998 a 2007

Empresa	Nº de anos classificada como rentabilidade superior	Rentabilidade para o acionista acumulada (%)
Souza Cruz	10	4.924
Lojas Americanas	9	1.527
Weg	10	1.255
Fosfértil	8	1.211
Gerdau Metal.	8	871
Fras-Le	7	775
Itausa	7	752
Metal Leve	7	696
Petrobras	7	681

No Brasil, os estudos mais recentes sobre previsão de falências (MARTINS, 2003; JANOT, 2001; ROCHA, 1999) utilizam o modelo *logit* e de análise de sobrevivência. Para minimizar a influencia das variáveis macroeconômicas, limitam-se a analisar e validar seus modelos em circunstâncias em que as condições macroeconômicas permanecem estáveis ou pelo menos sobre a mesma política econômica.

Sarkar (2001) introduz um modelo probabilístico empregando redes bayesianas para previsão de falências bancarias e Sun e Shenoy (2007) aperfeiçoam o modelo para empresas em geral. Modelos probabilísticos são baseados em um conjunto de informações probabilística devidamente codificadas que permitem calcular as probabilidades de ocorrência de quaisquer arranjos de informações contidas no modelo de acordo com os axiomas básicos da teoria de probabilidades (PEARL, 1988). Ou seja, dispondo de proposições autônomas pode-se calcular a probabilidade de ocorrer qualquer combinação dessas proposições. No caso do modelo probabilístico de previsão de falências mais comuns, constrói-se o modelo com o histórico de indicadores econômico-financeiros de empresas e fornecendo-se os dados de uma determinada empresa, a qualquer momento, obtém-se uma probabilidade de ocorrência da sua falência futura.

As Redes bayesianas consideram o que ocorreu no passado e incorporam o conhecimento anteriormente adquirido (aprendizado), bem como o grau de conhecimento sobre determinada variável. Como respostas, apontam as probabilidades de sucesso ou não de uma empresa. Essa metodologia ainda não foi aplicada no Brasil para a previsão de falências nem tão pouco para a previsão de rentabilidade futura.

O presente trabalho visa desenvolver uma ferramenta de apoio a decisão de investir empregando um classificador bayesiano que forneça a probabilidade de se obter uma rentabilidade futura superior para empresas de capital aberto brasileiras tendo como informação seus indicadores microeconômicos e não a previsão de falências das empresas apesar de ser baseado nesses. As variáveis classificadoras destes modelos são as probabilidades de uma empresa falir ou não o qual é o oposto do problema original do presente estudo. O objetivo não é determinar as probabilidades de uma empresa falir, mas sim, quais são suas probabilidades de alcançar uma rentabilidade superior em relação as demais. Pode-se dizer que é um modelo de previsão de falências “ao contrário”.

A construção dessa ferramenta será útil para estimar as chances de sucesso de um investidor ao avaliar uma determinada empresa e seus índices tendo como objetivo a rentabilidade superior. Almeja também contribuir para a difusão do emprego da técnica bayesiana na confecção de ferramentas de apoio a decisão.

O capitulo 2 faz uma breve revisão bibliográficas e são abordadas as bases teóricas para o desenvolvimento do trabalho mencionado os modelos estatísticos usuais de previsão e o modelo bayesiano, sua fundamentação probabilística, como é criado um modelo bayesiano, as Redes Bayesianas, os classificadores bayesianos e o conceito de *d-separado*. No capitulo 3 são apresentados a metodologia da pesquisa, os objetivos, o banco de dados, a seleção das variáveis, e os softwares utilizados.O capítulo 4 apresenta os modelos de previsão desenvolvidos, analiza as variáveis encontradas, valida com a regressão logística e compara-os entre si e analisa os resultados. Por fim, são apresentadas as conclusões finais com a limitação dos modelos e a recomendação para estudos futuros.

2 REVISÃO BIBLIOGRÁFICA E BASE TEÓRICA

Neste capítulo, será apresentada uma revisão dos trabalhos sobre previsões de falências, desenvolvidos até o momento, que são a base do desenvolvimento da ferramenta do presente estudo. Será feita menção ao trabalho de Sarkar e Sriram (2001), quem primeiro introduziu a abordagem bayesiana para a determinação da probabilidade de falências de empresas utilizando seus índices econômico-financeiros. Menciona o trabalho de Sun e Shenoy (2007) que apresenta uma metodologia estatística mais sólida para a determinação das variáveis explicativas e não somente opinião de especialistas como vinha sendo feito na maioria dos trabalhos até então e, em seguida, faz-se uma breve revisão de alguns modelos de previsão estatísticos incluindo o modelo bayesiano e a teoria de probabilidade subjacente. Como exemplo é apresentado um modelo de criação da ferramenta de previsão e, por fim, as Redes Bayesianas.

2.1 REVISÃO BIBLIOGRÁFICA

Os modelos de previsão de falências são estudados desde a década de 1930 e são várias as técnicas empregadas, tais como, análise discriminante (ALTMAN, 1968), redes neurais (ALFARO *et al.*, 2008; CELIK; KARATEPE, 2007; CHO *et al.*, 2009), regressão *logit* (SOHN; KIM, 2007; TSENG; LIN, 2005), sobrevivência proporcional (SHUMWAY, 2001), ou mesmo programação genética (MCKEE; LENSBERG, 2002; LENSBERG, *et al.*, 2006) dentre outros. Kumar e Ravi (2006) em sua revisão sobre esses estudos apresentam 74 trabalhos publicados, somente no período de 1968 até 2005. O número significativo de trabalhos sobre o tema demonstra, de antemão, sua importância.

Altman (1968) demonstrou que a falência das empresas poderiam ser prevista à partir da análise de seus índices econômico-financeiros e ajudou a ampliar os modelos de previsão de falências. Tais estudos podem ser úteis para os gestores na detecção de indícios de problemas financeiros futuros, para constatação da efetividade dos controles gerenciais e para fundamentar ações gerenciais corretivas ou preventivas. Para as instituições financeiras, na análise de crédito. Para o governo identificar a necessidade de intervir antecipadamente em uma instituição financeira e evitar a liquidação da mesma. Para auditores, evitar erros nos pareceres de auditorias além de subsidiar opiniões mais acertadas a respeito das previsões de uma determinada empresa (ZURITA, 2008; SARKAR; SRIRAM, 2001; KUMAR; RAVI, 2007).

Srakar e Sriram (2001) inovam ao introduzirem um modelo bayesiano para o problema e suas preocupações principais estavam na modelagem do processo de auditoria que pode ser implementado ao serem usados os mecanismos de atualização de crenças que a teoria probabilística de Bayes propicia. Esses autores usaram dois modelos, o classificador bayesiano simples (*Naive*) e o classificador composto e compararam as performances dos dois com o classificador de árvore de decisão. O classificador simples apresentou resultados satisfatórios. Outros estudos também apontam que o emprego de redes bayesianas como classificadores podem ser empregados em outros campos do conhecimento apresentado boa performance (ANDERSON *et al.*, 2004; SARKAR; SRIRAM, 2001) alem de permitir respostas em ambientes onde nem todos os dados estão disponíveis alem de não necessitarem do conhecimento prévio sobre a distribuição estatística das variáveis.

Sun e Shenoy (2007) usando uma amostra de 6932 empresas norte-americanas que não faliram e 890 que faliram no período se 1989-2002 dividiram essa amostra em dez e realizaram algumas simulações. Os autores listaram os vinte índices mais usados e usaram 2 tipos de classificador, o simples e o composto. O método de seleção dessas variáveis foi a verificação da correlação entre cada uma delas com o resultado esperado e aquelas cuja correlação fosse maior que 0,1 foram consideradas. Daí resultou numa classificação com oito variáveis principais. As demais variáveis que tiveram correlação com alguma variável principal foram consideradas como secundárias para a hipótese de não se ter algum valor estatístico da variável principal. Com base nos testes que fizeram, concluíram que as

diferenças de qualidade do classificador simples comparada com o classificador o composto não foram significativas. Outra constatação interessante foi quanto ao número de estados de cada variável que não deve ser maior que três estados por variável alem de que a substituição das variáveis discretas por variáveis continuas não aumenta o desempenho do modelo. Portanto, no presente estudo, será utilizado o classificador bayesiano simples, as variáveis serão discretas com no máximo três categorias

Nesse estudo foram incluídas duas variáveis que podem de certa forma capturar alguma informação do mercado que é a taxa de falência media de falências ocorridas nos últimos dois anos para o setor e a opinião do auditor.

Zurita (2008) faz uma comparação entre três modelos de previsão de falências, o de Risco de Crédito, o *Probit* e o de Sobrevivência e a novidade é a inclusão de variáveis macroeconômicas ao modelo possibilitando assim uma abordagem mais geral ao indicador. Outra consideração que faz em seu estudo é a consideração de outros eventos como importantes na determinação da variável dependente e inclui eventos além da falência pura e simples como a saída da empresa do mercado seja por incorporação ou venda. Tal flexibilização do modelo parece bastante razoável uma vez que o processo de falência em si é um caso extremo e, antes de ocorrer, outras providências provavelmente serão tomadas como a venda, fusão, reformulações internas, aumento de capitais etc.

Os modelos de previsão de falências também sofrem críticas contundentes de alguns pesquisadores que afirmam categoricamente que são “modelos que são desenvolvidos usando-se métodos e dados altamente questionáveis e que são impraticáveis” (NWOGUGU, 2007). Segundo este autor os riscos e as decisões são modelados de uma maneira mais correta ao se usar fatores quantitativos e qualitativos e diz entre outras considerações que o processo de falência é dinâmico e gradual e segue uma seqüência de eventos e que na maioria das vezes não se transforma em falência e sim na venda, fusão etc. Tal afirmação vem ao encontro do que considera Zurita (2008).

Para um melhor entendimento do modelo de previsão de falências com base nas redes bayesianas é necessário entender também as principais técnicas estatísticas de previsões tais como a Regressão Linear Múltipla, Regressão Logística e o Modelo de Sobrevivência bem como se constrói uma rede bayesiana, assunto dos próximos itens.

2.2 BASES TEÓRICAS

Para o entendimento do modelo de previsão é necessário o conhecimento acerca dos modelos de previsão estatísticos mais comuns e do modelo bayesiano.

2.2.1 Modelos de previsão com base estatística

Serão apresentados neste capítulo os modelos estatísticos mais comuns de previsão que são a Regressão Linear, a Regressão Logística, e os modelos de Sobrevivência. As Redes Bayesianas vêm a seguir como a forma de abordagem do problema de previsão de falência que foi escolhida pelo presente estudo.

2.2.1.1 Modelo de regressão linear

Procura-se na Regressão Linear Múltipla estimar o valor de uma variável y não observada, denominada de dependente, em função de outras variáveis x_n, denominadas de explicativas, observadas. Ou seja, dispondo de uma seqüência de dados $y_{1,}y_{2,}y_{3}.....y_{n}$ associados a esses valores de $x_{1,}x_{2,}x_{3,}...x_{n}$ ambos observados,

deseja-se encontrar valores não conhecidos de y para uma seqüência de x_n observados, ou seja:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ...\beta_n x_{nt} + \varepsilon_i$$

Onde ε_i representa o erro residual que ocorre entre a previsão e o acontecido, β_n são os parâmetros de estimação da regressão. x_{2i} representa , por exemplo, a i -ésima observação da variável explicativa x_2 e β_0 é o termo constante.

A equação de regressão é o exemplo mais conhecido de uma variável estatística entre as técnicas multivariadas (HAIR *et al.*, 2006).

2.2.1.2 Modelo de regressão logística

A regressão logística é apropriada quando se deseja como resposta uma variável binária, sim ou não, falência ou sucesso, doente ou sadio, etc. e muitos pesquisadores preferem-na por ser similar a regressão e possuem testes estatísticos diretos , tem a habilidade de incorporar efeitos não lineares e uma vasta gama de diagnósticos (HAIR *et al.* 2006). Para o calculo do coeficiente logístico deve-se comparar a probabilidade de um evento ocorrer com a de não ocorrer representado pela equação:

$$\frac{Probabilidade\ de\ um\ evento\ ocorrer}{Probabilidade\ de\ um\ evento\ não\ ocorrer} = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+...\beta_n x_n}$$

Com a transformação dessa equação em termos de logaritmo tem-se:

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... \beta_n x_n$$

Uma vez que o desejado é a probabilidade do evento ocorrer então p pode ser expresso pela equação conhecida como a função logística:

$$p = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... \beta_n x_n}}$$

Os cálculos para a obtenção dos coeficientes logísticos podem ser facilmente calculados com programas estatísticos como o SPSS e os testes de significâncias para eles são o da Razão da máxima verossimilhança, teste de *score* e pelo teste de Wald e são assintoticamente equivalentes e em muitas situação praticas, a aplicação de qualquer um deles leva às mesmas conclusões (LANDAU; EVERITT, 2004).

2.2.1.3 Modelo de sobrevivência

O principal atrativo dessa técnica é que ela fornece não só a probabilidade de sucesso ou fracasso, mas também uma estimativa do tempo para que isso ocorra dada pela equação:

$$S(t) = \Pr(T > t)$$

Essa equação pode ser estimada pelo quociente do numero de indivíduos que conseguiram sobreviver a um tempo determinado e o numero de indivíduos da amostra:

$$\hat{S}(t) = \frac{Numerodeindividuos\ com\ tempo\ de\ sobrevivencia > t}{Numero\ de\ indivíduos\ na\ amostra}$$

Define-se uma função de risco que expressa a probabilidade de um evento ocorrer num intervalo pequeno de tempo adiante dado que o evento não ocorreu até o presente momento. Por exemplo: A probabilidade de uma pessoa morrer aos 100 anos é pequena, pois normalmente as pessoas morrem antes deste tempo, entretanto a probabilidade desse indivíduo morrer daqui para frente é bastante grande (LANDAU; EVERITT, 2004). A representação dessa função seria :

$$h(t) = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots\beta_n x_n}$$

ou

$$\ln\left[h(t)\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\beta_n x_n$$

Essa representação somente é aplicável para funções que são constantes ao longo do tempo.Então criou-se a função de risco proporcional de Cox assim definida.

$$\ln\left[h(t)\right] = \ln\left[h_0(t)\right] + \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\beta_n x_n$$

Onde $\ln\left[h_0(t)\right]$ é denominada função de risco inicial para indivíduos cujas variáveis explanatórias são iguais a zero.

Esse modelo obriga que a razão do risco proporcional entre dois indivíduos se mantenha ao longo do tempo, ou seja, se uma empresa A tem 3 vezes mais chances de falir que uma empresa B no inicio da contagem, essa razão se mantém ao longo de toda a análise.Os parâmetros da regressão de Cox são estimados pela máxima verossimilhança e podem ser obtidos com o uso de softwares estatísticos.

2.2.1.4 Modelo bayesiano

O modelo bayesiano é determinado através da presunção de independência entre as variáveis sendo que a probabilidade condicional é dada pela fórmula:

$$\text{Prob}(A \mid B) = \frac{P(A,B)}{P(B)} = \frac{\text{Probabilidade de que } A \textbf{ e } B \text{ aconteçam simultaneamente}}{\text{Probabilidade de que } B \text{ aconteça}}$$

Onde $P(A,B) = P(A \cap B)$

No caso do classificador simples a representação gráfica do indicador é o da Figura 1.

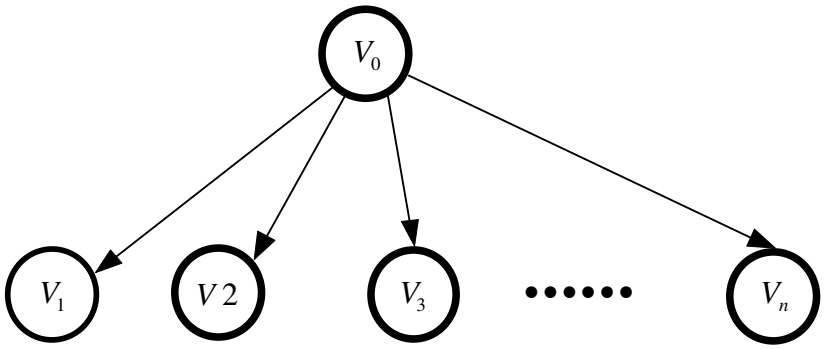


Figura 1 - Classificador bayesiano simples V_o = variável classificadora e $V_1 ... V_n$ = variáveis explicativas

A resposta dada pela variável classificadora não precisa ser necessariamente dicotômica embora estudos revelem que a quantidade ideal não deva passar de 3 (SUN; SHENOY, 2007). Para se entender melhor como funciona o classificador se faz necessária uma revisão da teoria da probabilidade e sua aplicação na confecção do classificador bayesiano o que será apresentado a seguir.

2.2.1.4.1 A formulação bayesiana da probabilidade

No formalismo bayesiano a medida da crença se mede através de três axiomas da teoria da probabilidade:

1- $0 \leq P(\ E\ ) \leq 1$

2- $P(\ evento\ certo\ ) = 1$

3- $P(\ A\ ou\ B\ ) = P(\ A\ ) + P(\ B\ )$ se A e B são mutuamente excludentes

O terceiro axioma estabelece que a crença atribuída a qualquer conjunto de eventos é a soma das crenças atribuídas aos elementos que não se interceptam. Assim, cada evento A pode ser interpretado como a união da probabilidade conjunta dos eventos (AeB) e, $(\ A\ e\ nãoB\ )$ e as probabilidades associadas a eles são dadas por:

$$P(\ A\ ) = P(\ A\ e\ B\ ) + P(\ A\ e\ nãoB\ )$$

Generalizando, se $B_i, i = 1,2,3...n$ é um conjunto mutuamente excludente, (também chamado de partição ou uma variável), então $P(A)$ pode ser encontrada de $P(AeB_i)$ usando-se a soma:

$$P(\ A\ ) = \sum_i P(\ A, B_i\ )$$

Esta formula pode ser ilustrada pela Figura 2:

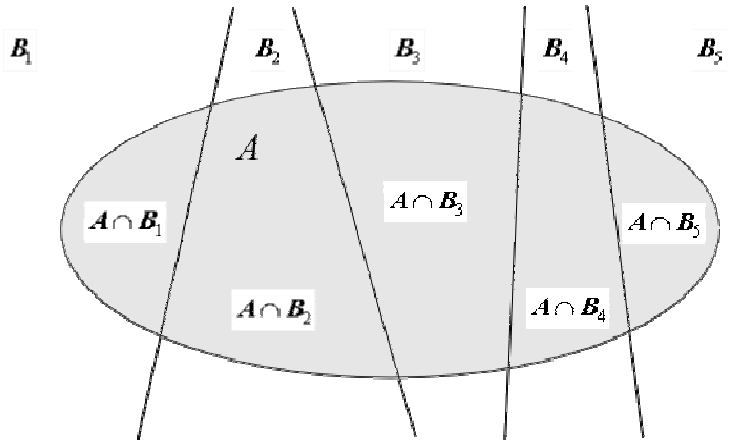


Figura 2 - A soma das intersecções forma o todo

Por exemplo, a probabilidade de A= Arremesso de dois dados serem iguais pode ser escrita somando-se a probabilidade dos eventos conjuntos (A, B_i) = 1,2...,6, onde B_i é a proposição "O primeiro arremesso apresentou o resultado i".

$$P(A) = \sum_i P(A, B_i) = \text{ 6 x 1/36 = 1/6}$$

A conseqüência é que a proposição mais a sua negativa devem ser atribuídas a probabilidade 1 ou,

$$P(A) + P(\sim A) = 1$$

Pois uma das alternativas dever ser verdadeira.

As expressões básicas no formalismo bayesiano são afirmações sobre probabilidades condicionais isto é, $P(A \mid B)$. Determina-se a crença em A na condição de B ser conhecido com absoluta certeza. Se $P(A \mid B) = P(A)$ é dito que A e B são independentes se $P(A \mid B, C) = P(A \mid C)$ é dito que A e B são condicionalmente independentes dado C.

A probabilidade condicional é dada pela formula:

$$P(A \mid B) = \frac{P(A,B)}{P(B)}$$

ou

$$P(A,B) = P(A|B)P(B)$$

Uma vez que $P(A,B) = P(B,A)$ então:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Sabendo-se que os eventos são mutuamente excludentes tem-se então que:

$$P(A) = \sum_i P(A,B_i) = \sum_i P(A \mid B_i)P(B_i)$$

O centro da técnica Bayesiana é a formula da inversão definida por:

$$P(H|e) = \frac{P(e|H)P(H)}{P(e)}$$

Ou seja, a crença de que a hipótese H aconteça dado que a evidencia e aconteceu pode ser calculada ao se multiplicar a crença inicial $P(H)$ de pela verossimilhança $P(e \mid H)$ de que a evidencia se materializará em H seja verdade. $P(H \mid e)$ é as vezes chamada de probabilidade posterior e $P(H)$ é chamada de probabilidade *a priore*. O denominador $P(e)$ raramente entra em consideração por ser uma constante *normalizadora* $P(e) = P(e \mid H)P(H) + (Pe \mid nãoH)P(nãoH)$ que pode ser computada por ser necessário que a soma de $P(H \mid e)$ mais $P(nãoH \mid e)$ resultem na unidade (PEARL, 1988).

2.3 A CRIAÇAO DE UM MODELO BAYESIANO

A construção deste modelo, a primeira vista, parece intuitiva, pois bastaria analisar o histórico de empresas que faliram e com base em seus indicadores financeiros estabelecer uma probabilidade para cada combinação dos valores destes indicadores. Como exemplo deseja-se construir um modelo com 4 variáveis dicotômicas (bom, ruim) como Ativo total (At), Lucro Liquido (Ll), Liquidez geral (Lg), Passivo circulante (Pc) e dispondo de pelo menos 10 casos onde estes indicadores tenham o mesmo atributo estabelecer a probabilidade de falência para uma futura empresa assim, se temos 10 casos onde $At = b,\ LL = r,\ Lg = r,\ Pc = r$, e neste 10 casos 40 % das empresas faliram então a probabilidade de se ter uma empresa falir e que apresente estes mesmos atributos é de 40 %. O que ocorre, entretanto, é que não se dispõe de tantos dados assim, pois no caso de 4 variáveis são de 2^4 = 16 as possíveis combinações destes indicadores e para se obter uma probabilidade ao menos razoável há a necessidade de pelo menos 10 casos para cada combinação de indicadores. Sendo assim, seriam necessários um número superior a 160 casos. Se forem considerados 10 indicadores as combinações possíveis destes seriam de 2^{10} = 1024 o que inviabilizariam estimativas razoáveis para o modelo hipotético.

Para contornar esse problema assume-se a independência entre as variáveis e assim viabiliza-se a construção do modelo, ou seja, no caso do modelo hipotético anterior, At é independente de Ll que é independente da Lg que é independente do Pc. Então, se considera que dado que a empresa faliu, o fato de se ter um At bom ou ruim nada influencia o fato de se ter um Pc bom ou ruim. Assim, a atribuição das probabilidades dos indicadores pode ser calculada independentemente, ou seja, dispondo de 10 casos e destes, 3 apresentam $Pc = ruim$ e a empresa faliu, independentemente de se avaliar os outros indicadores a $P(Pc = ruim \mid faliu)$ é de 30% e com base nisso e no teorema de Bayes é possível chegar a um resultado.

Como exemplo ilustrativo, considerando-se a independência entre as variáveis e o teorema de Bayes em uma amostra de 30 empresas onde 10 delas faliram e 20 delas não. Serão considerados os indicadores de Liquidez geral e de Passivo circulante conforme o Quadro 1. Nesta tabela, o número da segunda coluna e terceira linha corresponde ao numero de empresas (3) cujo indicador Liquidez geral estava bom e mesmo assim faliu, o numero da quarta linha e da quinta coluna (11) representa a quantidade de empresas cujo passivo circulante era ruim e que não faliram.

Tabela 2 - Exemplo ilustrativo de uma amostra com 30 empresas e alguns de seus índices

	Faliu		**Não faliu**		**Total**
	Bom	**Ruim**	**Bom**	**Ruim**	
Liquidez geral	3	7	15	5	30
Passivo circulante	8	2	9	11	30
Total	**10**		**20**		**30**

Deseja-se obter a probabilidade de uma empresa falir dadas as condições de sua liquidez geral ser boa e o passivo circulante ser ruim. Esse cálculo é possível com a utilização do teorema de Bayes e as probabilidades condicionais onde:

$$\text{Prob(A|B)}=\frac{\text{Prob(A,B)}}{\text{Prob(B)}}=\frac{\text{Probabilidade de que A e B aconteçam simultaneamente}}{\text{Probabilidade de que B aconteça}}$$

$$P\left(S = Faliu \mid Lg = Bom, Pc = Ruim\right) = \frac{P(S = Faliu, Lg = Bom, Pc = Ruim)}{P(Lg = Bom, Pc = Ruim)}$$

Sabendo-se que:

$$P(S = Faliu, Lg = Bom, Pc = Ruim) =$$

$$P\left(Lg = Bom \mid S = Faliu\right) * P\left(Pc = Ruim \mid S = Faliu\right) * P\left(S = Faliu\right)$$

$$P\left(Lg = Bom \mid S = Faliu\right) = \frac{3}{10}$$

$$P\left(Pc = Ruim \mid S = Faliu\right) = \frac{2}{10}$$

$$P\left(S = Faliu\right) = \frac{10}{30}$$

Assim, $P(S = Faliu, Lg = Bom, Pc = Ruim) = 0,0200$

$$P(Lg = Bom, Pc = Ruim) =$$

$$P\left(Lg = Bom \mid S = Faliu\right) * P\left(Pc = Ruim \mid S = Faliu\right) * P\left(S = Faliu\right) +$$
$$P\left(Lg = Bom \mid S = NãoFaliu\right) * P\left(Pc = Ruim \mid S = NãoFaliu\right) * P\left(S = NãoFaliu\right)$$

$$= \frac{3}{10} * \frac{2}{10} * \frac{10}{30} + \frac{15}{20} * \frac{11}{20} * \frac{20}{30} = 0,0200 + 0,2750 = 0,2950$$

Portanto :

$$P\left(S = Faliu \mid Lg = Bom, Pc = Ruim\right) = \frac{0,0200}{0,2950} = 0,0678$$

Como se pode ver, a quantidade de dados necessária para a elaboração do classificador de falências desejado fica bastante reduzida, sendo assim, o exemplo anterior é passível de ser estendido para o emprego de mais variáveis com o mesmo tamanho de amostra o que, por si só, já é uma justificativa convincente para a utilização desse método.

2.4 REDES BAYESIANAS

O emprego atual de redes bayesianos abrange um amplo espectro, seja para a previsões de preços futuros do petróleo (ABRAMSON; FINIZZA, 1995), previsões meteorológicas (ABRAMSON *et al.*, 1996), substituto de operadores de algoritmos genéticos (PELIKAN, 2005; KOBLIHA *et al.*, 2006; LI; AICKELIN, 2003) e muitos outros que possivelmente virão. Deve-se portanto entender o que são e como são criadas.

2.4.1 Definições

Redes bayesianas são modelos gráficos representados por um conjunto de variáveis e um conjunto de arcos que ligam algumas dessas variáveis e que apresentam tabelas de probabilidades condicionais entre as variáveis interligadas. A variável que dá origem a um arco é dita variável pai e a variável cujo arco aponta é dita filha. Assim adotando a definição de (JENSEN, 2007) Redes Bayesianas são:

- Um conjunto de variáveis e um conjunto de arcos direcionados que unem algumas ou todas as variáveis;
- Cada variável possui um conjunto de estados finito e mutuamente excludente;
- As variáveis juntamente com os arcos direcionadas formam um grafo chamado de Grafo direcionado e este é dito acíclico se não existe nenhum caminho direcionado $A_1 \to A_2 \to ...A_n$ em que $A_1 = A_n$;
- A cada variável A com parentes $B_1,...B_n$, é adicionada uma tabela de probabilidades condicionais $P(A \mid B_1,...B_n)$.

Quando usadas em conjunto com técnicas estatísticas o modelo apresenta vários atributos para a análise de dados que são:

1- Devido ao modelo apresentar probabilidades condicionais entre as variáveis, este lida com situações em que alguns valores para as variáveis estão ausentes;
2- A rede Bayesiana pode ser usada para se aprender relações causais entre as variáveis e pode ser usada para se incrementar o aprendizado sobre o domínio do problema e prever conseqüências ao se intervir no problema;
3- Com o modelo possuindo a relação semântica causal e probabilística é ideal para se combinar conhecimento anterior (que freqüentemente vem na forma causal) com os dados presentes;
4- A estatística Bayesiana em conjunto com a RB oferece uma abordagem eficiente para se evitar um super aprendizado (HECHERMAN, 1966).

2.4.2 Criando uma rede bayesiana

Roteiro básico para se criar uma RB

1) Criar as variáveis que representam o modelo
   Dado um problema, é necessário transformar-lo em uma linguagem que seja possível o emprego do formalismo matemático para a resolução desse problema assim, no caso da determinação de empresas lucrativas ou não condicionada a ao capital de giro, uma variável seria o Lucro e a outra o capital de giro;
2) Definir para cada variável um conjunto de estados
   No exemplo anterior os estado possíveis dessas variáveis poderiam ser: para o lucro, alto ou baixo; para o capital de giro, baixo, médio ou alto;
3) Estabelecer as relações de dependência causal entre as variáveis criando arcos entre pais e filhos. Para o mesmo exemplo, a presunção de que o capital de giro afeta o lucro das empresas, deverá então haver um relação causal entre as duas variáveis no sentido de capital de giro para lucro ou

vice-versa e mesmo uma relação sem causa e efeito entre elas (ver Figura 3);

4) Avaliar as probabilidades, *a priori*, fornecendo ao modelo valores probabilísticos para cada variável. Conhecendo-se as variáveis, os estados e as relações causais entre eles, é necessário estabelecer qual a probabilidade de um evento ocorrer dado que o outro evento ocorreu assim, com base em constatações anteriores, estabelecer as probabilidades anteriores.

Exemplo de RB

Um investidor deseja aplicar seus recursos em uma empresa e os dois dados de que dispõe são: a taxa de retorno desta empresa que foi classificada como superior em 40% dos casos e os dados do endividamento desta empresa. Sabendo-se que empresas que normalmente tem taxas de retorno superiores apresentam 10% de seu endividamento classificado como baixo, 40% classificado como médio e 50% classificado como alto e que das empresas que obtiveram taxa de retorno baixa 30% delas apresentavam um endividamento baixo, 40% médio e 30% alto. Quais as chances que o investidor teria em aplicar seu dinheiro nesta empresa dado o grau de endividamento desta empresa?

Primeiro passo: Determinação das variáveis que seriam Sucesso e Endividamento;

Segundo passo: Estados para as variáveis: Para Sucesso seria superior e baixo, para Endividamento, baixo, médio e alto;

Terceiro passo: O estabelecimento da relação causal é demonstrado na Figura 3 onde o primeiro nó representa a variável Sucesso e o segundo nó a variável Endividamento;

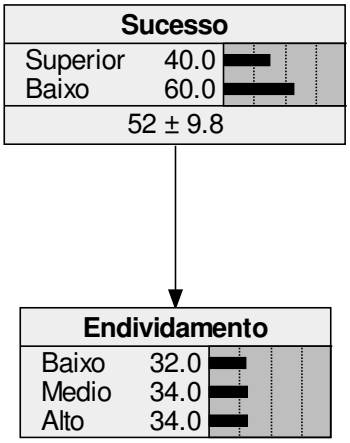


Figura 3 - Relação causal entre as variáveis

Quarto passo: Os valores a priori foram inseridos na rede conforme a Tab. 3;

a) Para o primeiro nó que representa as probabilidades de sucesso de um novo empreendimento qualquer;

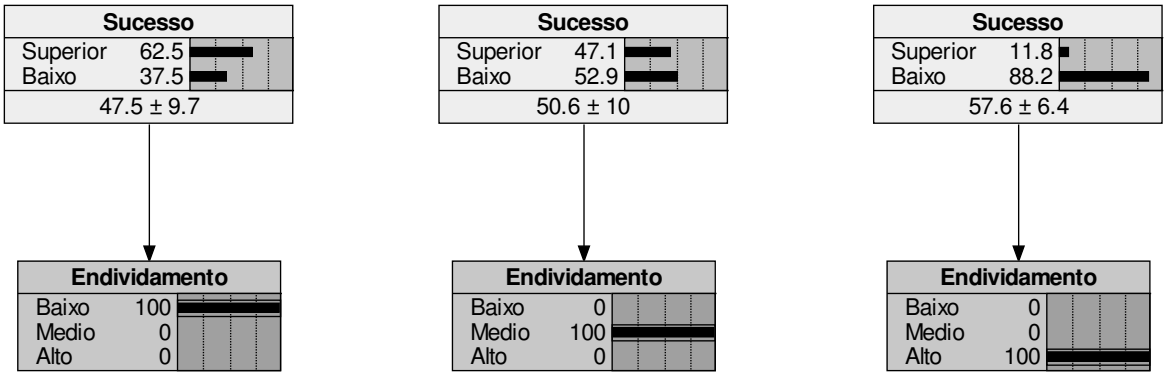
Sucesso	Chances(%)
Superior	40
Baixo	60

b) Para o nó que representa o endividamento: $P(End \mid Suc)$

Tabela 3 - Valores das probabilidades condicionais

Endiv.Superior	Superior (%)	Baixo (%)
Baixo	50	20
Médio	40	30
Alto	10	50
∑	**100**	**100**

Com os dados de que dispõe o investidor os resultados para as três hipóteses de endividamento seriam a da Figura 4:



(a) Baixo Endividamento **(b) Endividamento médio** **(c) Endividamento alta**

Figura 4 - Três hipóteses de Endividamento e suas conseqüências para a estimativa de sucesso

Para que o investidor tenha mais certeza sobre suas aplicações outra variável poderia ser incluída no modelo tal como a variável Margem Bruta (MB). Os valores hipotéticos seriam: das empresas que obtiveram sucesso superior, 60% delas apresentavam Margem Bruta superior, 30% delas MB médio e 10% delas MB inferior. Das empresas que tiveram desempenho inferior, 60 % delas apresentavam MB inferior, 30% MB médio e 10% delas MB superior. A nova Rede construída é o da Figura 5 já com as três variáveis.

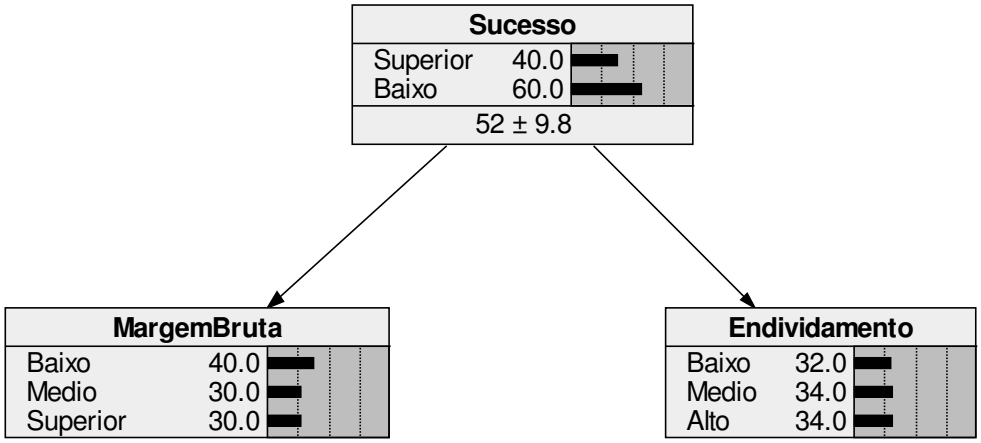


Figura 5 - Nova rede construída com a inclusão de uma nova variável Margem Bruta

Tabela 4 - Cálculo da Probabilidade de Sucesso Dado o Grau de endividamento e a Margem Bruta

Variável Margem Bruta	Variável Endividamento	Probabilidade de sucesso (%) sup.
Baixo	Baixo	21,7
	Médio	12,9
	Alto	2,17
Média	Baixo	62,5
	Médio	47,1
	Alto	11,8
Alta	Baixo	90,9
	Médio	84,2
	Alto	44,4

Com três variáveis os cálculos já não ficam tão simples pois são 9 os casos possíveis cujos resultados estão na Tabela 4.

Assim, os modelos desenvolvidos no presente trabalho seguem essa mesma metodologia e o próximo passo será o entendimento sobre os classificadores bayesianos que é o centro da ferramenta desenvolvida.

2.4.3 Classificador bayesiano

Dado um conjunto de variáveis $A = \{A_1, A_2, ...A_n\}$ denominados atributos ou características e uma Variável *C* denominada classificadora e que possam ser atribuídas estados, um classificador é uma função de *A* que dá uma resposta quanto a *C*, como exemplo, se for classificar uma animal quanto a sua espécie dado suas características quanto ao formato do corpo, mamífero ou ovíparo, coberto de pelos, penas, escamas, número de patas etc, sua espécie poderia ser determinada .

Um caso *completo* é quando se dispõe de todos os atributos para se determinar uma classe, o que não é muito comum, e um caso *consistente* é quando casos completos com os mesmos atributos resultem na mesma classe (JENSEN, 2007).

As Redes Bayesianas podem ser usadas como classificadores desde que possuam uma única variável de resposta, como no caso do exemplo do investidor que avalia a possibilidade de se investir em um novo empreendimento sendo a única variável de resposta para o investidor investir ou não no empreendimento dado que dispõem de algumas informações. A variável classificadora seria então $Invertir\{sim, não\}$ e as variáveis atributos $A = \{Margembruta, Endividamento\}$.

Um classificador Bayesiano simples é o que tem a variável classificadora como único pai de todos as variáveis atributos o que indica que a estrutura é fixa. O único trabalho que o pesquisador terá será o de determinar os parâmetros para essa Rede Bayesiana.

Existem também outros classificadores tais como o TAN no qual as variáveis atributos possuem pelo menos uma outra variável atributo como pai conforme a Figura 6. Outro classificador bayesiano é o da Arvore classificadora que tem os nós internos variáveis de atributos. Aos arcos são atribuídos valores para os atributos e às folhas, valores da classificação como mostrado na Figura 7.

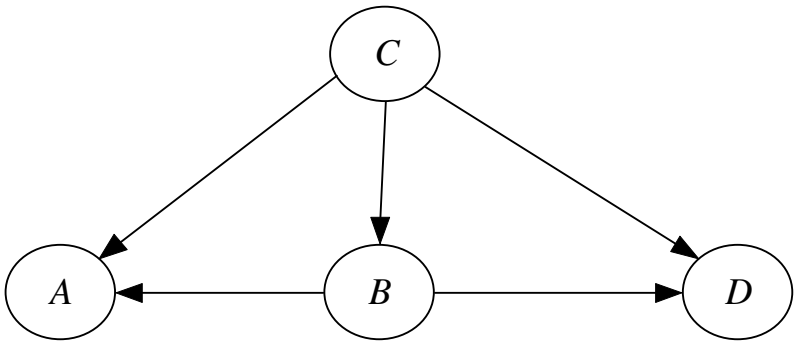


Figura 6 - Exemplo de classificador TAN

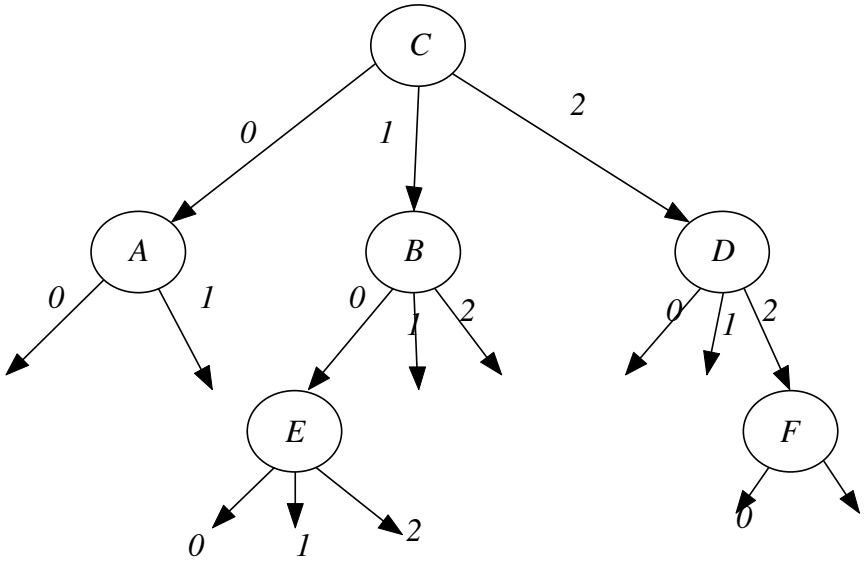


Figura 7 - Arvore classificadora

O presente estudo tratará somente do classificador bayesiano simples por já ter sido comprovada sua eficiência com estruturas mais complexas nas pesquisas realizadas por Sarkar e Sriram (2001) e Sun e Shenoy (2007).

2.4.4 Parâmetros de uma rede bayesiana

Definindo uma RB como $R = (E, \Theta)$ onde E representa a estrutura da rede e Θ representa os parâmetros da rede, são quatro as situações que se depara na construção da mesma:

1- Sabe-se a estrutura e os parâmetros;
2- Sabe-se a estrutura e não se sabe os parâmetros;
3- Não se sabe a estrutura, mas sabe-se os parâmetros;
4- Não se sabe a estrutura nem os parâmetros.

No estudo da determinação de um classificador Bayesiano simples a condição que se encontrará é a segunda uma vez que a estrutura já é conhecida pela própria definição do classificador. O que se desejará definir é qual a variável classificadora e quais as variáveis explicativas a empregar, no primeiro caso isso foi definido de antemão e no segundo será o objeto principal do presente estudo.

2.4.5 O conceito de *d-separado*

O conceito de *d-separado* é bastante útil para a simplificação dos cálculos das tabelas de probabilidades condicionais quando da execução dos algoritmos de determinação das mesmas. O presente estudo não pretende abordar a maneira como os cálculos são feitos, mas o conceito de *d-separado* ajuda a entender a simplificação que é adotada no problema foco que é a independência entre as variáveis explicativas (PEARL, 1988).

São três os tipos de conexões entre as variáveis, a serial, a divergente e a convergente. Na conexão serial da Figura 8, *A* tem influencia sobre *B* que tem influencia sobre *C*. Sabendo-se o estado de *C*, o julgamento do estado de *A* será alterado passando-se pelo julgamento do estado de *B.* Em outras palavras, evidências de *C* alteram a crença de *B* que por sua vez altera a crença em *A*. Sabendo-se o estado de *B*, a crença no estado de *A* se altera mas nesse caso qualquer evidência em *C* não se transmite a *A* pois o estado de *B* já é conhecido, assim concluí-se que evidencias de *C* se transmitem até *A* se não houverem evidencias de *B*. Diz-se dessa condição que *A* e *C* são d-separados dado *B,* ou seja, se tornam independentes dado *B*.

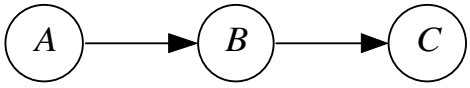


Figura 8 - Conexão serial

Como exemplo, sendo a variável *C* a variável que representa o fornecimento de energia elétrica pelas hidroelétricas que pode ser constante ou intermitente. A variável nível dos reservatórios (*B)* pode ser alto ou baixo A variável A é o índice pluviométrico do período (*A*) e pode ser alto ou baixo. Dado que está acontecendo uma interrupção no fornecimento de energia elétrica, pode-se supor que o índice pluviométrico do período foi baixo. Entretanto, ao se saber que o nível dos reservatórios está baixo tem-se a certeza que o nível pluviométrico foi baixo portanto se estiver ocorrendo interrupção ou não no fornecimento de energia a crença sobre o índice pluviométrico não se altera ao passo que se o conhecimento sobre o nível do reservatório não fosse conhecido a crença sobre o incide pluviométrico se alteraria. Portanto, o conhecimento de *C* não altera o conhecimento sobre *A* dado que se sabe sobre *B*. Assim, *C* é *d-separado* de *A* dado *B*

Na conexão divergente como o da Figura 9, se não houver nenhuma evidencia sobre *A* , tendo evidências em qualquer um de seus filhos altera-se a crença nos demais mas uma vez conhecido do estado de *A* , qualquer evidencia sobre um filho não altera a crença nos demais pois *A* já está determinado.

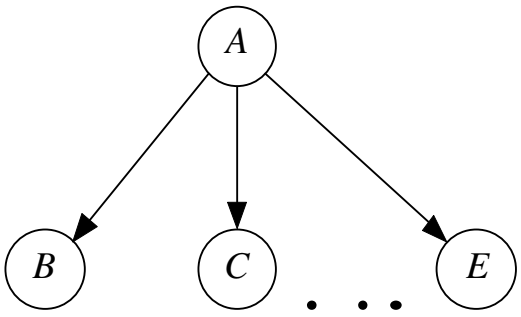


Figura 9 - Conexão Divergente

Como exemplo: Seja *A* uma variável que classifica um animal como sendo um pássaro ou não. *B* representa a presença de asas e *C* a presença de penas. Sabendo-se que o animal possui asas, a crença de que possui penas pode ser alterada, entretanto sabendo-se que é um pássaro, a constatação de que possui asas não altera a crença de que possui penas ou não. Assim, *B* é *d-separado* de *C* dado que o estado de *A* é conhecido.

No caso da conexão convergente (Figura 10), o raciocínio opera de maneira oposta, ou seja, o conhecimento sobre a variável *A* é que permite a conecção entre as variáveis pais.

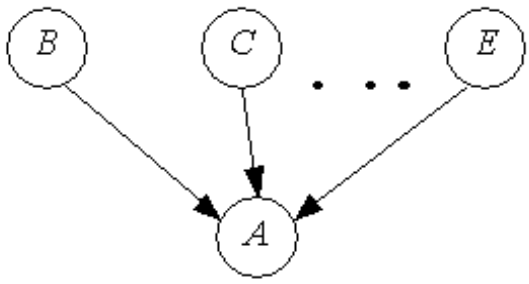


Figura 10 - Conexão convergente

Como exemplo: *A* representa dor de cabeça, *B* representa gripe e *C* stress nervoso. Sabendo-se que a dor de cabeça está presente, o conhecimento de que se tem stress pode alterar a crença de que se tem gripe e a falta de conhecimento sobre a dor de cabeça não altera as suposições sobre gripe e stress.Assim *B* e *C* são d-separados caso *A* não seja conhecido .

No caso do classificador bayesiano do presente trabalho sobre Rentabilidade a conexão é a divergente. Sabendo-se que uma empresa é rentável, bloqueia as conexões entres as variáveis filhas . Não conhecendo o estado da rentabilidade o conhecimento sobre o índice de endividamento pode alterar a crença sobre o giro de estoque e a margem bruta mas se o fato de ser lucrativa é conhecido, a crença sobre as variávies filhas não se altera.Pode-se dizer portanto que as variáveis filhas são d-separadas dado que a variável classificadora é conhecida.

3 METODOLOGIA DA PESQUISA

Neste capítulo serão apresentados os objetivos do estudo o metodo empregado para se chegar ao resultado final, bem como, o banco de dados utilizado, as variávies consideradas e os softwares empregados.

3.1 OBJETIVOS

3.1.1 Objetivo geral

O objetivo do presente trabalho é desenvolver uma ferramenta de apoio à decisão aplicando redes bayesianas, que indique a probabilidade de uma empresa alcançar, no ano seguinte, um retorno ao acionista superior às demais baseando-se no conhecimento de seus índices econômico-financeiros no presente.

3.1.2 Objetivos secundários

Para a construção dessa ferramenta, são necessários a busca dos seguintes objetivos secundários:

1-Determinação das variáveis explicativas;
2-Definição da Rede Bayesiana;
3-Validação do Modelo Bayesiano.

3.2 METODO DA PESQUISA

Os dados relativos aos balanços foram obtidos junto a empresa PARTNER que oference a seus clintes dados atualizados, do período de 1994 até 2007, de todas as empresas que são obrigadas a publicarem seus balanços (ver Anexo A). Tais dados foram tratados conforme a metodologia desenvolvidad por Sun e Shenoby (2007) , que seleciona quais as variáveis explicativas a serem escolhidas, bem como, transforma-as em categóricas tendo como variável classificadora a Rentabilidade sobre o Patrimonio liquido das empresas no ano seguinte. Foram desenvolvidos dois modelos bayesianos. No primeiro a variável classificadora é dicotômica e no segundo esta possui três estados. Para efeito de validação do modelo Bayesiano, este foi comparado com o modelo Logit que é uma metodologia frequentemente empregada em modelos de previsão de falencias.

A sequencia do trabalho está representada no fluxograma da Figura 11 e consiste no seguinte:

1- O Banco de Dados, foi adquiridos da Empresa Sabe-Partner que vende aos interessados os balanços de todas as empresas que são obrigadas a publica-los.Tal banco de dados vem sob a forma de tabelas e gráficos e para processsá-los é necessário a transformação desses dados no formato de planilha Excell;
2- Os modelos tradicionais de previsão de falências normalmente são apresentados como dois estados para a variável classificadora entretanto no intuito de averiguar se esse seria a melhor abordagem optou-se por desenvolver uma segundo modelo que dispusesse de mais um estado intermediário de classifção assim, todos os índices economicos-financeiros das empresas foram agrupados em dois ou tres grupos;
3- Ao se separar os grupos calculou-se as médias dos mesmos e verificou-se se eram significativamente diferentes entre si, pois não sendo assim o indice não apresentaria nenhum potencial explicativo, portanto foi descartado;
4- O passo seguinte foi a verificação das correlações desses indices cujas médias apresentaram diferenças significativas com a variável

classificadora que foi escolhida como sendo a Rentabilidade ao Acionista mas do ano seguinte a da publicação dos indices do balanço. A essa variável foi dado o nome de RAAdv;

5- Algumas das variáveis explicativas apresentaram correlações parciais entre si, portanto houve a necessidade de se averiguar se os pares não eram redundantes, ou seja, caso a presença de um provocasse no seu par uma correlação não significativa nem significante com a variável classificadora uma das variáveis deveria ser excluída.
6- Após a determinação de quais as variáveis a serem consideradas é necessária que essas sejam categóricas e ,para tanto, com base no histograma de cada uma delas, essa foram divididas em dois ou três grupos conforme o caso;
7- Os dados foram divididos por ano;
8- Para cada ano, foi calculada a tabela de probabilidade condicional da Rede Bayesiana e com os dados do ano seguinte, calculada a taxa de sucesso de cada ano;
9- Em paralelo, foi rodado um modelo Logit para a validação do modelo Bayesiano;
10- O resultado do primeiro modelo foi comparado com o modelo Logit;
11- Os dois modelos foram comparados entre si
12- As variáveis foram analisadas para cada caso;
13- O trabalho foi encerrado.

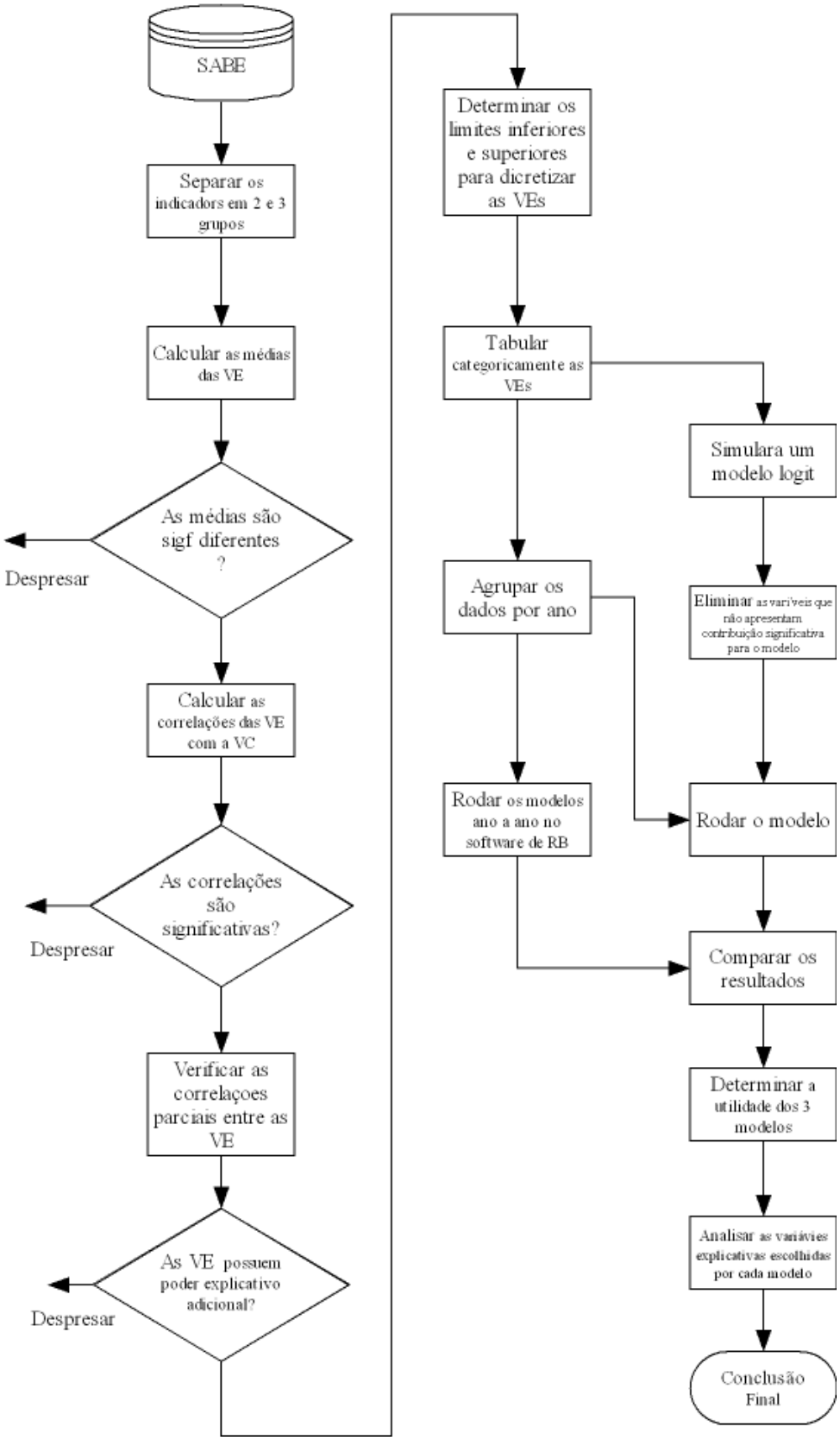


Figura 11 - Fluxograma do trabalho desenvolvido. VE= variável explicativa, VC= Variável classificadora

A seguir serão apresentados mais detalhes sobre o Banco de Dados, seleçao da variável classificadora e explicativas e os softwares utilizados.

3.2.1 Banco de dados

Como fonte de dados foi utilizada o SABE-PARTNER (Sistema de Analise de Balanços Empresarial, Versão Aula) que é um sistema que contém os dados de balanços de todas as empresas que negociam ações na Bolsa de Valores no período de 1994 até 2007 perfazendo um total de 9756 registros. Tal sistema dispõe de uma série de índices contábeis para análise financeira e que doravante serão chamados de variáveis e foram esses os empregados no presente trabalho (ver Quadro 1).

Índices Econômico-Financeiros

Variável	Abreviatura	Descrição	Formula
Lucro Líquido ($ Mil)	LL	Resultado da empresa no exercício fiscal.	Lucro Operacional + Receitas Não Operacionais – Despesas Não Operacionais – Provisão para IR + Reversão IR – Participação no Lucro +/ - Participação dos Minoritários
Endividamento Total sobre Ativo Total (%)	ETAT	Indica a participação do capital de terceiros de curto e longo prazos no Ativo Total Médio da empresa.	((Passivo Circulante + Exigível Longo Prazo + Duplicatas Descontadas) / Ativo Total) x 100 ETPL= Endividamento Total sobre o Patrimônio Líquido (%) ((Passivo Circulante + Exigível Longo Prazo + Duplicatas Descontadas) / Patrimônio Líquido) x 100
Endividamento Total sobre o Patrimônio Líquido (%)	ETPL	Mede a proporção dos recursos de terceiros em relação aos recursos próprios existentes na empresa.	((Passivo Circulante + Exigível Longo Prazo + Duplicatas Descontadas) / Patrimônio Líquido) x 100

Continua...

... continuação.

Endividamento Oneroso sobre Ativo Total (%)	EOAT	Indica a participação do capital de terceiros de curto e longo prazos no Ativo Total Médio da empresa.	(Empréstimos e Financiamentos Totais de Curto Prazo+Empréstimos e Financiamentos Totais de Longo Prazo + Debêntures de Curto Prazo+Debêntures de Longo+Empresas Coligadas e Controladas de Curto Prazo+Empresas Coligadas e Controladas de Longo Prazo+Duplicatas Descontadas)/ (Ativo Total) x 100
Endividamento Oneroso sobre Patrimônio Líquido (%)	EOPL	Indica a participação das fontes externas de financiamento (Recursos Externos) em relação ao capital próprio.	(Empréstimos e Financiamentos Totais de Curto Prazo+Empréstimos e Financiamentos Totais de Longo Prazo + Debêntures de Curto Prazo+Debêntures de Longo+Empresas Coligadas e Controladas de Curto Prazo+Empresas Coligadas e Controladas de Longo Prazo+Duplicatas Descontadas)/ (Patrimônio Líquido) x 100
Endividamento Oneroso de Curto Prazo sobre Oneroso Total (%)	EOCPOT	Indica a porcentagem do volume da dívida onerosa de curto prazo em relação a dívida onerosa total (curto e longo prazo).	(Empréstimos e Financiamentos Totais de Curto Prazo+ Debêntures de Curto Prazo+ +Empresas Coligadas e Controladas de Curto Prazo+ Duplicatas Descontadas) / (Empréstimos e Financiamentos Totais de Curto Prazo+Empréstimos e Financiamentos de Longo Prazo + Debêntures de Curto Prazo+Debêntures de Longo+Empresas Coligadas e Controladas de Curto Prazo+Empresas Coligadas e Controladas de Longo Prazo+Duplicatas Descontadas) x 100
Composição do Endividamento (%)	CE	Indica qual o percentual de obrigações a curto prazo em relação às obrigações totais.	(Passivo Circulante + Duplicatas Descontadas) / (Passivo Circulante + Exigível Longo Prazo + Duplicatas Descontadas) x 100
Imobilização do Patrimônio Líquido (%)	IPL	Identifica a participação do Patrimônio Líquido em relação ao volume total investido no Ativo Permanente.	(Ativo Permanente / Patrimônio Líquido) x 100
Imobilizados dos Recursos Permanentes (%)	IRP	Identifica a participação dos recursos não correntes em relação ao volume total investido no Ativo Permanente.	(Ativo Permanente) / (Exigível Longo Prazo + Patrimônio Líquido) x 100

Continua...

... continuação.

Imobilização do capital próprio	ICP	Identifica a participação do Patrimônio Líquido em relação ao volume total investido no Ativo Permanente.	Fórmula: (ATIVO PERMANENTE / PATRIMÔNIO LÍQUIDO) * 100
Imobilização dos recursos permanentes	IRP	Identifica a participação dos recursos não correntes em relação ao volume total investido no Ativo Permanente.	Descrição: Identifica a participação dos recursos não correntes em relação ao volume total investido no Ativo Permanente. Fórmula: (Ativo Permanente / (Exigível a Longo Prazo + Patrimônio Líquido)) * 100
Margem Bruta (%)	MB	Identifica o desempenho dos custos de produção. Dado que o lucro bruto é obtido da diferença entre as vendas líquidas e o custo dos produtos vendidos ou das mercadorias vendidas, um aumento deste índice significa uma melhor eficiência produtiva da empresa.	(Lucro Bruto / Receita Operacional Líquida) x 100
Margem Operacional (%)	MO	: Identifica o desempenho operacional da empresa computando o resultado financeiro mais o resultado de equivalência patrimonial.	(Lucro Operacional / Receita Operacional Líquida) x 100
Margem Líquida (%)	ML	: Representa a percentagem de cada real de venda que permaneceu na empresa após o pagamento de todas as despesas inclusive financeiras e imposto de renda.	(Lucro Líquido / Receita Operacional Líquida) x 100
Margem EBIT (%)	MEBIT	Representa a percentagem de cada unidade monetária da recita líquida que resultou em EBIT.	EBIT / Receita Operacional Líquida) x 100
Margem EBITDA (%)	MEBITDA	Representa a percentagem de cada unidade monetária da recita líquida que resultou em EBITDA.	(EBITDA / Receita Operacional Líquida) x 100

Continua...

... continuação.

Giro do Ativo (X)	GA	Indica quanto a empresa vendeu para cada R$ 1,00 de investimento total (Ativo Econômico), ou seja, identifica o número de vezes que o Ativo Total da empresa girou em determinado período em função das vendas realizadas.	(Receita Operacional Líquida / Ativo Total) x 100
Retorno dos Acionistas (%)	RA	Também conhecido como ROE (Return on Equity), indica o retorno dos recursos próprios investidos na empresa. É expresso pela relação entre os resultado líquido obtidos em determinado período e capital próprio empregado.	(Lucro Líquido / Patrimônio Líquido) x 100
Rentabilidade do Ativo Total (%)	RAT	Também conhecido como ROI (Return on Investment) ou ROA (Return on Assets), indica a lucratividade operacional que a empresa propicia em relação ao total de investimentos.	(Lucro Líquido / Ativo Total) x 100
Patrimônio Liquido	PL	É o valor que os proprietários têm aplicado	Contas do patrimônio líquido têm saldos credores, divide-se em: Capital social; Reservas de capital; Reservas de reavaliação, Reservas de lucros; e Lucros/Prejuízos acumulados.
Retorno do Investimento Total (%)	RI	Também conhecido como ROI (Return on Investment) ou ROA (Return on Assets), indica a lucratividade operacional que a empresa propicia em relação ao total de investimentos.	(Lucro Operacional / Ativo Total) x 100
Termômetro Financeiro (x)	TI		(Saldo de Tesouraria / Necessidade de Capital de Giro)

Quadro 1 - Indicadores financeiros considerados nos modelos

3.2.2 Variável classificadora

A variável classificadora é a que responde à questão no modelo. O propósito do presente estudo é auxiliar um investidor que quer saber onde aplicar seu dinheiro tendo como informação os índices econômicos financeiros de uma empresa. Sob esse ponto de vista, mas sem excluir a possibilidade da adoção de qualquer outro indicador, dois índices de lucratividade foram candidatos, o retorno do acionista (RA) que é o lucro líquido da empresa dividido pelo patrimônio liquido e o retorno do investimento total que é o lucro operacional dividido pelo ativo total (RIT). Neste ultimo, como os ativos representados no balanço de uma empresa nem sempre correspondem a realidade e muitas vezes estão subavaliados (ZURITA, 2008) optou-se pelo Retorno ao acionista futuro (RAAdv) como a variável classificadora.

Foram testados dois modelos de classificação: no primeiro a variável classificadora foi dividida em dois estados (baixa, superior) e no segundo em três estados (baixa, mediana, superior).

Para transformar as variáveis contínuas em discretas adotou-se o critério de Sun e Shenoy (2007) embasados em estudos anteriores que constatatam ser este o melhor critério .Para o segundo modelo empresas cujo desempenho fossem inferiores a marca correspondente a 18,5 % da distribuição dos desempenho das empresas num determinado ano foi atribuído a classificação baixa, empresas cujo RA fossem inferiores a marca de 81,5 % (18,5%+63%) e superiores a marca anterior foram classificadas como desempenho mediano e como superior foram classificadas as empresas que apresentarem desempenho superior a 81,5 %. Para o primeiro modelo, empresas que tiveram desempenho classificados abaixo de 81,5% foram classificadas como baixo e acima dessa marca como superior.

3.2.3 Seleção das variáveis explicativas

A metodologia empregada para a determinação das variáveis explicativas foi baseada no estudo de Sun e Shenoy (2007) com a inclusão de uma etapa inicial que foi a determinação das média dos índices entre os grupos. Como premissas básicas da construção do modelo, serão consideradas, que deverão haver três estados para cada variável no máximo, ou seja, ao tornar uma variável continua em discreta, serão considerados no máximo três estados para elas, uma vez que, ao aumentar o numero de estados, o modelo perde em previsibilidade. O modelo será o classificador bayesiano simples, pois o modelo composto não apresentou aumento da previsibilidade. Baseado nessas premissas, as etapas para a determinação das variáveis explicativas foram as seguintes:

1. Determinar as diferenças significativas entre as médias dos indicadores financeiros dos 2 (ou 3) grupos de interesse (medianas e superiores; baixas medianas, superiores);
2. Determinar a correlação entre as principais variáveis escolhidas com a variável classificadora e verificar a correlação de Pearson entre elas. Se for maior que 0,10 serão incluídas no modelo, se não, excluídas. No presente estudo optou-se por adotar a marca de 0.05 por estar-se buscando correlações mais sutis do que a falência;
3. Estabelecer a correlação parcial entre as variáveis que tenham correlação significativa entre si e verificar a intensidade entre elas se <0,01 excluir. Optou-se por considerar todas as correlações que fossem significantes ao nível de 0,05;
4. Tornar as variáveis continua em discretas dividindo-as em, no máximo, 3 estados conforme sua probabilidade acumulada ou seja, para valores que estejam abaixo do percentil de 18,5 % =baixo, para valores entre 18,5% e 81,5% =médio e acima de 81,5% bom.

3.2.4 Softwares utilizados

Para a construção e teste dos classificadores, foram utilizados dois softwares específicos para o trabalho com redes Bayesianas, o Genie2 e o Netica. O primeiro é um software, de distribuição livre, desenvolvido pelo Laboratório de Sistemas de Decisão da Universidade de Pittsburgh, tem a vantagem de dispor de algoritmos de estimativa de parâmetros além da estimativa da estrutura, é de fácil manejo, dispõe de ferramentas de calculo estatísticos básicos além de permitir a fácil exportação e importação de dados de outros programas como e Excel e o Netica.

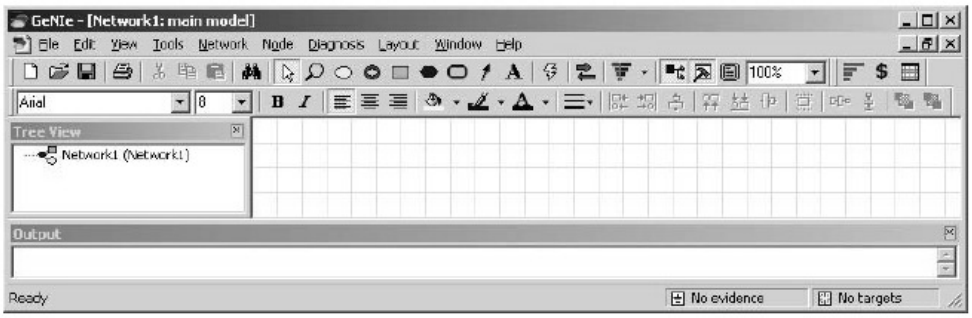


Figura 12 - Tela principal do Genie2

O Software, Nética é distribuído por Norsys Software Corp. e é pago entretanto dispõe de um modulo limitado a 15 nós o que não compromete o presente estudo e a vantagem em relação ao anterior, é a disponibilização de ferramentas para a avaliação das Redes Bayesianas como a matriz de confusão, acerácea, qualidade, etc.

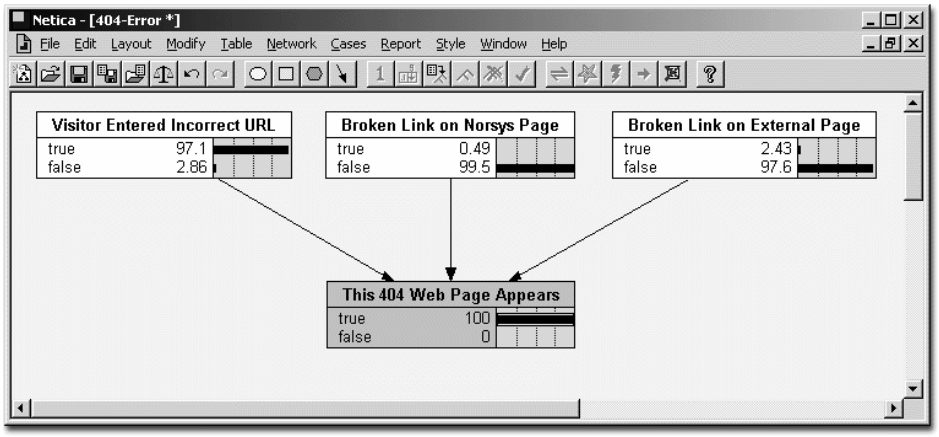


Figura 13 - Tela principal do Nética

Assim, o Genie2 foi usado na construção inicial da rede e o Nética para a avaliação do classificador.

4 MODELOS BAYESIANOS DE PREVISÃO DE RENTABILIDADE

Normalmente, nos modelos de previsão de falências somente dois estados são apresentados: faliu ou não faliu. Mas no caso da previsão da rentabilidade futura, além do modelo dicotômico foi criado mais um que introduz um estado intermediário na variável classificadora, pois ao se apostar em uma rentabilidade superior e se obter uma rentabilidade mediana isso poderia representar um ganho significativo em relação a média das rentabilidades.O primeiro modelo não só apresenta uma média de acertos superior ao segundo como também uma dispersão menor mas ao se realizar a comparação entre eles utilizando-se um modelo de utilidade o segundo modelo apresentou um desempenho superior. O segundo modelo com base na metodologia aplicada seleciona mais variáveis do que o primeiro, mas mesmo assim não tem mais regularidade.

4.1 O PRIMEIRO MODELO

O propósito do presente estudo não é investigar nem discutir detalhadamente os indicadores financeiros utilizados pelos analistas contábeis, mas desenvolver uma ferramenta para descobrir relações subjacentes entre eles, portanto, quanto aos índices, serão apresentadas somente as descrições dos mesmos que foram transcritas conforme consta no banco de dados do SABE que já foram apresentados no Quadro 1.

Os dados de todas as empresas foram considerados para o período de 1996 até 2007. Todos os índices econômicos financeiros disponíveis que apresentam razões ou proporções foram considerados e comparados com a variável classificadora que é a Rentabilidade do Acionista do ano posterior (RAAdv). Os dados de todas as empresas foram tabulados e seus indicadores foram posicionados seqüencialmente por ano. As empresas que não dispunham das informações completas foram excluídas, tais como, as que dispunham dos índices em 2006, mas que não apresentaram balanço em 2007, ou empresas que iniciaram

a publicação do balanço em 1999 e que não dispunham do índice em 1998 também foram excluídos da análise.

Ao analisar o histograma de cada variável foi constatado que havia uma grande concentração de valores nos extremos e isso dificultava a obtenção dos limites de corte para a transformação em variáveis categóricas (ver Gráfico 1) e para evitar distorções, uma vez que o SABE atribui valor -100 a todas as empresas que apresentam Patrimônio líquido negativo e algumas poucas empresas apresentavam RA muito superior a 100. Assim, somente quando dos cálculos das variáveis e para a determinação dos limites inferiores e superiores, foram considerados dos dados onde -100<RA<100 assim, as distorções foram minoradas tornando possível a análise.Entretanto,todas as empresas foram consideradas quando da discretizaçao das variávies.

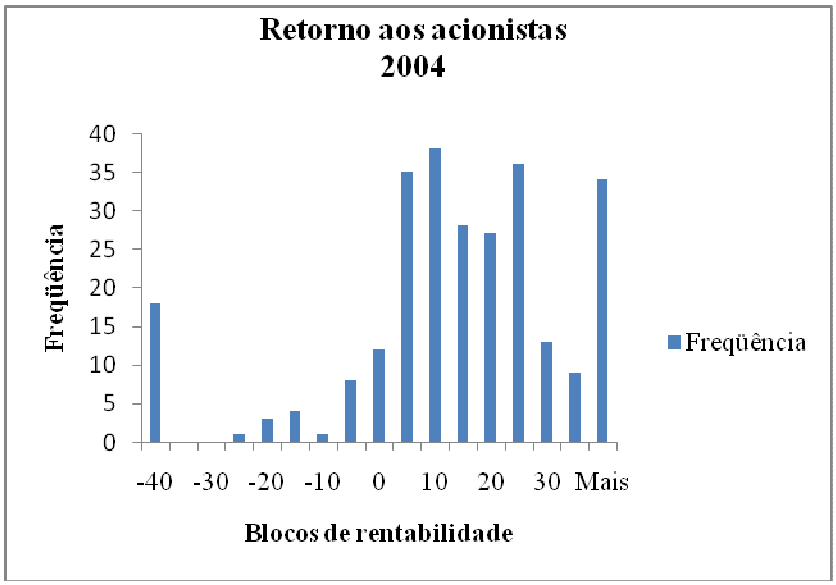


Gráfico 1 - Histograma da rentabilidade aos acionistas das empresas no ano de 2004. Um grande número de empresa nos extremos

Restaram então, 3944 conjuntos de dados para a análise das médias entre os grupos e da correlação entre a variável classificadora (RAAdv) e os indicadores financeiros disponíveis. No primeiro modelo, as variáveis que apresentaram diferenças significantes entre as médias dos grupos foram CE, EOPL, EOAT, ETPL, GA, MB conforme pode-se ver no Anexo A .

As variáveis que tivessem médias diferentes foram separadas e verificou-se a correlação entre elas com a variável classificadora. A variável EOAT foi eliminada pois não apresentou correlação significativa com a RAAdv (Tabela 5).

Tabela 5 - Modelo 1. Correlação entre as variáveis candidatas após a primeira seleção. A variável EOAT é eliminada por não apresentar correlação significante com a variável RAAdv

		CE	EOPL	EOAT	ETPL	GA	MB	RAAdv
CE	Correlação de Pearson	1	-,073**	-,062**	-,063**	,225**	-,127**	,091**
	Sig. (2-caudas)		.000	.000	.000	.000	.000	.000
	N	3702	3702	3702	3702	3702	3702	3702
EOPL	Correlação de Pearson	-,073**	1	.001	,947**	-.006	.006	,050**
	Sig. (2-caudas)	.000		.955	.000	.734	.696	.002
	N	3702	3702	3702	3702	3702	3702	3702
EOAT	Correlação de Pearson	-,062**	.001	1	-.001	-.022	-.025	-.017
	Sig. (2-caudas)	.00	5		.955	.175	.129	.288
	N	3702	3702	3702	3702	3702	3702	3702
ETPL	Correlação de Pearson	-,063**	,947**	-.001	1	.029	.009	,050**
	Sig. (2-caudas)	.000	.000	.955		.081	.564	.003
	N	3702	3702	3702	3702	3702	3702	3702
GA	Correlação de Pearson	,225**	-.006	-.022	.029	1	-.004	,095**
	Sig. (2-caudas)	.000	.734	.175	.081		.805	.000
	N	3702	3702	3702	3702	3702	3702	3702
MB	Correlação de Pearson	-,127**	.006	-.025	.009	-.004	1	,048**
	Sig. (2-caudas)	.000	.696	.129	.564	.805		.004
	N	3702	3702	3702	3702	3702	3702	3702
RAAdv	Correlação de Pearson	,091**	,050**	-.017	,050**	,095**	,048**	1
	Sig. (2-caudas)	.000	.002	.288	.003	.000	.004	
	N	3702	3702	3702	3702	3702	3702	3702

**. Correlação significante ao nível de 0.01 (2-caudas).

Todas as correlações significantes ao nível de 0,05 e superiores a 0,05 foram considerados, com exceção da variável MB que apresentou uma correlação igual a 0,048 mas por ser potencialmente explicativa foi mantida no modelo.

Uma vez que as variáveis possuem correlação entre elas (Figura 14) foi necessário a verificação da correlação parcial entre cada uma delas para observar o possível poder explicativo adicional de cada uma em relação a outra conforme mostra a Tabela 6.

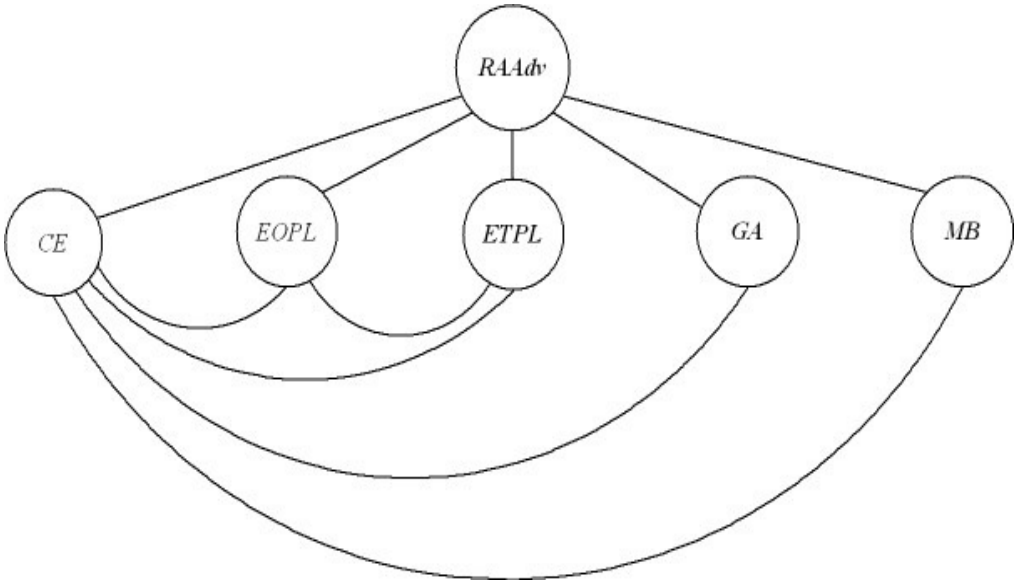


Figura 14 - Correlações entre as variáveis

A correlação parcial mede a relação entre duas variáveis desprezando-se o efeito de uma terceira variável em um modelo de regressão múltipla. Considerando-se uma matriz de correlações com três variáveis, *A, B, C* onde as correlações são : r_{AB}, r_{AC}, r_{BC} a correlação parcial entre *A* e *B* fixando-se *C* é dado por:

$$r_{ABC} = \frac{r_{AB} - r_{AC} r_{BC}}{\sqrt{\left(1 - r_{AC}^2\right)\left(1 - r_{BC}^2\right)}}$$

Tabela 6 - Primeiro modelo:Correlação parcial entre as variáveis explicativas candidatas

Variável considerada c/RA	Variável de Controle	Correlação Parcial/Sig	Variáveis consideradas	Variável de Controle	Correlação Parcial/Sig
CE	EOPL	0,095/000	EOPL	CE	0,057/0,001
ETPL	CE	0,056/0,001	CE	ETPL	0,094/0,000
EOPL	ETPL	0,01/0,557	ETPL	EOPL	0,007/0,681
GA	CE	0,077/0,000	CE	GA	0,072/0,000
ETPL	GA	0,047/0,004	GA	ETPL	0,093/0,000
MB	CE	0,060/0,000	CE	MB	0,098/0,000
ETPL	MB	0,049/0,003	MB	ETPL	0,048/0,004

A variável EOPL foi eliminada por não apresentar poder explicativo significante em relação a ETPL e esta já havia sido escolhida quando comparada com CE. Mais uma vez , MB seria candidata a exclusão por apresentar correlação parcial de 0,048 com ETPL e esta também poderia ser excluída por apresentar correlação parcial de 0,047 com GA mas como foram significantes as correlações parciais, decidiu-se manter-las no modelo.

Tabela 7 - Primeiro modelo: Limites para a classificação das variáveis explicativas em categóricas com base nos histogramas de cada uma delas

		Limites Modelo I							
Ano	**Tamanho da Amostra**	**Limites CE**		**Limites ETPL**		**GA**	**Limites MB**		**Mediana RA**
		Inferior	**Superior**	**Inferior**	**Superior**	**Mediana**	**Inferior**	**Superior**	**Divisor**
1996	257	34.0894	92.3688	21.8988	152.6731	0.4800	8.9296	49.1158	12.593
1997	293	28.8956	93.4788	18.8742	156.7960	0.4600	5.3465	53.9098	14.31
1998	342	28.1357	98.6855	16.7951	205.2604	0.3600	0	56.5366	15.61
1999	335	29.1376	97.8084	13.8704	215.8802	0.3600	0	48.7561	13.44
2000	319	27.5323	96.2998	14.4761	231.2701	0.3700	0	49.4797	15.55
2001	294	27.3657	94.5734	15.9484	232.4090	0.3700	0	51.8309	16.88
2002	289	27.7602	94.2509	9.7885	259.8361	0.3400	0	46.4177	19.16
2003	289	30.1189	92.0839	12.7586	277.1350	0.4000	0	45.0334	23.85
2004	273	26.1994	93.7887	9.6807	273.5474	0.4500	0	45.1538	25.57
2005	277	25.5079	93.6607	10.3676	250.3184	0.4200	0	46.8361	25.81
2006	290	25.1916	97.128	7.3044	215.7565	0.3000	0	52.6577	22.47

Para as variáveis selecionadas foram calculados os limites superiores e inferiores por ano para a transformação em variáveis discretas (Tabela 8). Quando da verificação do histograma de cada variável constatou-se que o GA apresentou uma característica de que mais de 18,5 % das empresas ficaram abaixo desse limites ou mesmo iguais a zero, portanto, optou-se por condicioná-lo em dois estados sendo o parâmetro de separação sua mediana anual.

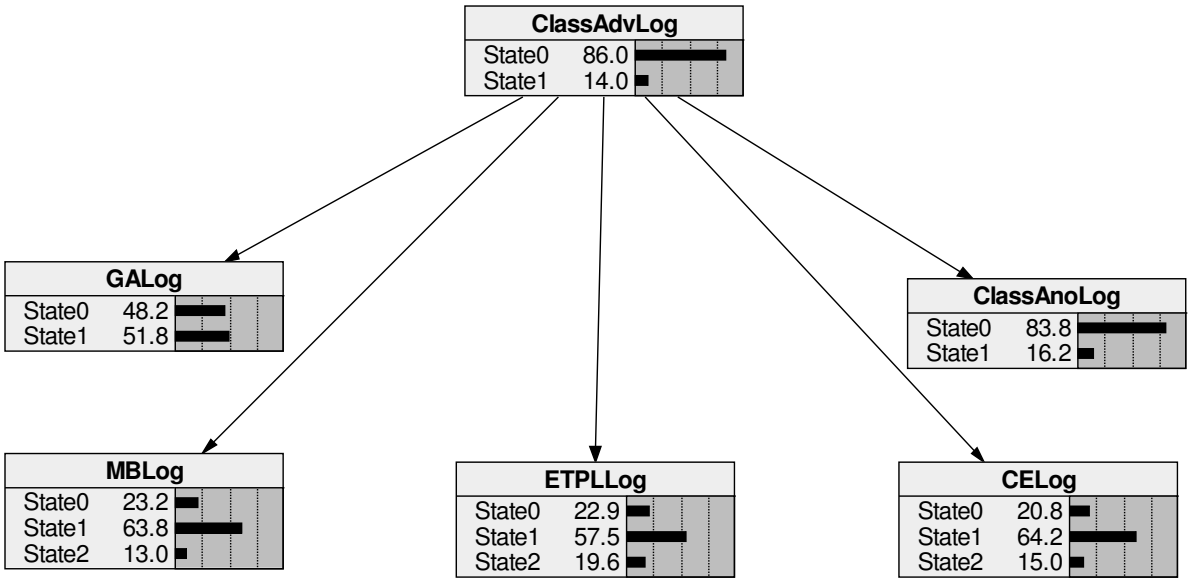


Figura 15 - Primeiro modelos: Classificador da Rentabilidade futura das empresas

Foi incluído também a variável, ClassAno, atribuído o valor 1 para a empresa que no ano presente se posicionou como superior e valor 0 para empresas que se posicionaram abaixo de 85,1%. O modelo obtido foi o representado na Figura 15 e a tabela de probabilidades condicionais estão na Tabela 8.

Tabela 8 - Probabilidades condicionais das variáveis no modelo 1. Dados de todos os anos reunidos

Variável	Estado	Variável Classificadora	
		Estado 0 (Baixa) (%)	Estado 1 (Superior)(%)
MB	0	27,78	20,34
	1	58,17	61,66
	2	14,06	18,60
GA	0	50,82	41,85
	1	49,18	58,15
ETPL	0	27,83	17,53
	1	55,61	60,72
	2	16,56	21,75
CE	0	21,94	13,46
	1	64,50	70,00
	2	13,56	16,75
ClassAno	0	90,39	46,08
	1	9,61	53,92

4.1.1 Análise das principais variáveis explicativas I

Para efeito de análise das variáveis explicativas, apesar de este não ser o foco principal do presente estudo, foram considerados todos os dados disponíveis, pois dessa forma pode-se avaliar o potencial explicativo médio de cada variável e não somente as de um determinado ano. Para tanto foi criada a Tabela 9 que representa o potencial explicativo de cada variável ao ser mantido inalterado o grau de crença das demais variáveis variando-se o grau de crença da variável em questão, ou seja, no caso da variável ETPL com 3 estados foi considerado o estado 0 como fixo e como resultado obteve-se a probabilidade da empresa apresentar um RAAdv superior de 10,03%; alterando-se o estado de ETPL para 1 o resultado foi de 16,20% e alterando para o estado 2 o resultado foi de 18,87%. Para as variáveis ClassAno e GA que só apresentam dois estados obteve-se uma probabilidade para cada um deles e assim sucessivamente.

Tabela 9 - Variação das probabilidades encontradas na variável classificadora ao se variar o grau de crença da variável explicativa segundo seus estados possíveis mantendo-se inalterados os graus de crença das demais variáveis explicativas

Grau de impacto	Variável Explicativa	Nº de estados	Baixo (%)	Mediano (%)	Superior (%)
1º	ClassAno	2	8,28	--	49,82
2º	ETPL	3	10,03	16,20	18,87
3º	CE	3	9,80	16,10	17,95
4º	MB	3	11,48	15,08	18,48
5º	GA	2	12,72	--	17,30

A ordem, segundo o grau de impacto encontrado, para cada variável é o que aparece na primeira coluna da Tabela 9, ordem essa determinada pelo software Netica ao se solicitar a sensibilidade da rede. A classificação da empresa no ano corrente (ClassAno) é a que tem maior impacto na previsão da rentabilidade futura o que não causa nenhuma estranheza uma vez que é de se esperar que empresas com bom desempenho no presente ano são as mais prováveis de serem classificadas com bom desempenho no ano seguinte. Como segunda classificada figura a variável ETPL que é o Endividamento total da empresa divididos pelo patrimônio liquido que indica que empresas que trabalham bem com dinheiro de

terceiros são as que são as mais propensas a terem uma RA superior no ano seguinte. Tal constatação é reforçada pela terceira colocada CE que é a composição do Endividamento que indica quanto da divida da empresa deverá ser paga a curto prazo ou seja, as dividas de curto prazo dividido pelo total da dívida. A quarta colocada é a margem bruta (MB) o que dispensa comentários pois é lógico supor que quanto maior a margem que uma empresa consegue melhores são as chances de obter uma boa rentabilidade . Por último, então, figura o Giro do ativo que é a Receita Operacional Liquida dividida pelo Ativo Total e sua interpretação é de quanto mais a empresa consegue girar seu ativo melhor sua rentabilidade que parece intuitivo.

4.2 O SEGUNDO MODELO

No segundo modelo, procurou-se atribuir um estado intermediário à variável classificadora divididos conforme o critério anteriormente descrito (se abaixo do percentil de 18,5 % = baixo (0), para valores entre 18,5% e 81,5% = médio (1) e acima de 81,5% superior (2).

Todas as variáveis em questão foram divididas em três grupos e, conforme essa separação, calculadas as médias dos mesmos. Essas, por sua vez, foram testadas as significâncias das diferenças com o teste LSD e as variáveis candidatas foram: CE, ETPL, GA, ICO, IRP e MB (Anexo B). Houve correlação significativa entre CE e ETPL, GA, IRP, MB; ETPL e IRP, ICP, e MB; ETPL e IRP (Figura 16). TF não apresentou correlação significativa com RAAdv (0,011/0,519) portanto foi eliminado.

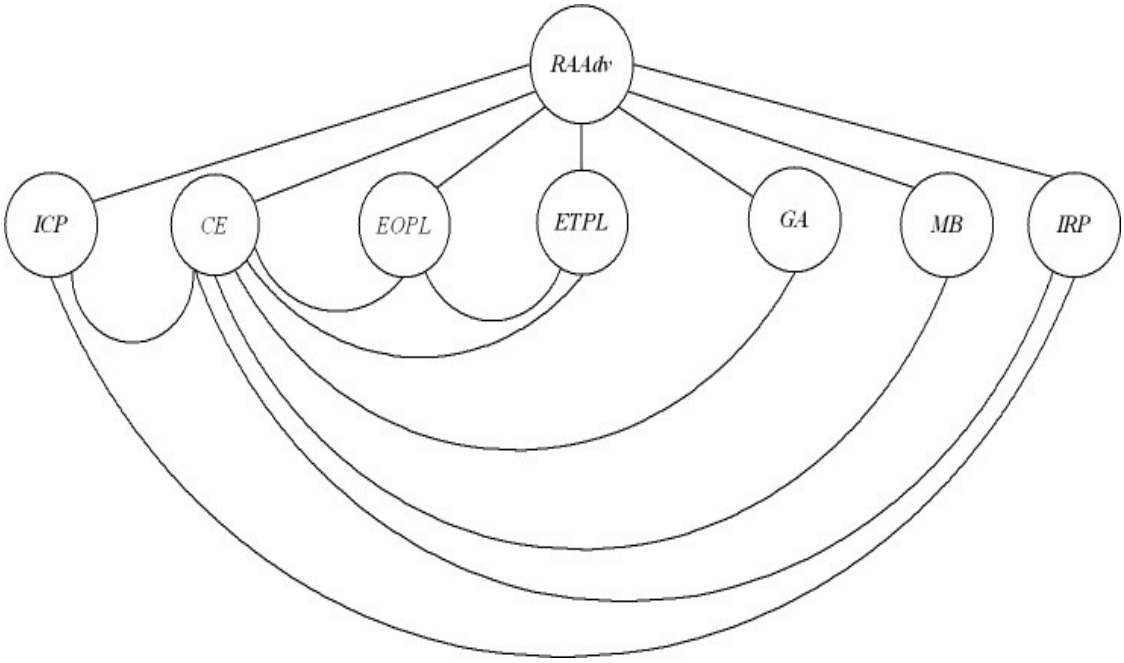


Figura 16 - Correlações entre as variáveis do segundo modelo

Calculadas as correlações parciais entre os pares de variáveis que ainda não o foram no modelo anterior conclui-se que nenhum índice seria eliminado (Tabela 10), pois, o potencial explicativo adicional de cada uma delas controlando-se seu respectivo par foi significativo. MB e IRP não apresentarem diferenças significativas entre duas de suas médias mas mesmo assim foram mantidos por apresentarem correlações significativas ao nível de 0,01 com a variável classificadora.

Tabela 10 - Segundo modelo: Correlações parciais entre as variáveis candidatas

Variável considerada c/RA	Variável de Controle	Correlação Parcial/Sig	Variáveis consideradas	Variável de Controle	Correlação Parcial/Sig
ICP	IRP	0,05/0,00	IRP	ICP	-0,076/0,00
ICP	CE	0,053/0,001	CE	ICP	0,095/0,00
CE	IRP	0,094/0,00	IRP	CE	-0,077/0,00
CE	MB	0,098/0,00	MB	CE	0,060/0,00
CE	GA	-0,072/0,00	GA	CE	0,077/0,00

Os limites de classificação para cada variável estão descritos na Tabela 11 e o segundo modelo tem a forma representada na Figura 17 e as tabelas de probabilidades condicionais das variáveis explicativas estão na Tabela 12. Cabe lembrar que essa ultima foi calculada com todos os dados disponíveis e não representa a que foi utilizada ano a ano.

Tabela 11 - Segundo modelo: Limites para as variáveis ETPL E IRP anexadas ao primeiro modelo

		ICP		IRP	
Ano	**Amostra**	**Lim.Inf.**	**Lim.Sup**	**Lim.Inf.**	**Lim.Sup**
1996	257	73.1979	147.8930	58.9967	103.0407
1997	293	72.0449	162.0727	61.3991	101.9469
1998	342	75.6769	175.0861	61.3188	102.3235
1999	335	69.5124	191.9647	60.7954	101.8573
2000	319	71.0837	185.2288	56.2251	101.3022
2001	294	67.5878	175.7108	50.1297	100.7155
2002	289	59.4188	194.1939	50.5018	101.5810
2003	289	63.2819	188.7832	51.2642	100.6531
2004	273	51.7742	180.8952	44.844	99.4728
2005	277	55.4498	182.6331	42.8795	99.093
2006	290	46.2856	167.1923	41.104	99.8896

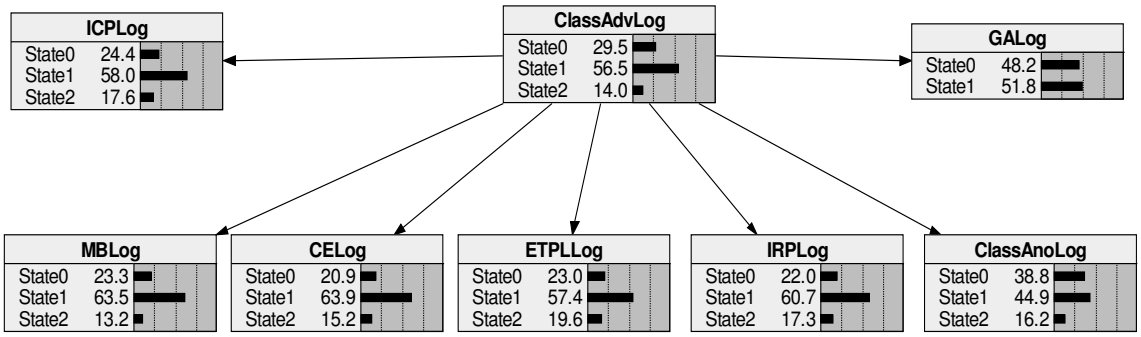


Figura 17 - Segundo modelo com três estados para a variável classificadora

Tabela 12 - Probabilidades condicionais das variáveis no modelo 2. Dados de todos os anos reunidos

Variável Estado		**Variável Classificadora (RAAdv)**		
		Estado 0 (Baixa) (%)	**Estado 1 (Mediana)(%)**	**Estado 2 (Superior) (%)**
MB	0	44,87	61,02	27080
	1	44,11	31,48	55,89
	2	11,12	7,50	16,31
GA	0	61,74	73,72	47,30
	1	38,27	26,28	52,70
ETPL	0	52,07	58,53	25,25
	1	26,00	35,79	55,04
	2	21,93	5,68	19,71
CE	0	43,09	56,95	21,56
	1	48,69	34,99	63,26
	2	8,22	8,06	15,18
ClassAno	0	58,13	19,75	13,62
	1	38,43	73,67	37,59
	2	13,62	37,59	48,79
ICP	0	50,61	58,93	25,67
	1	26,64	35,28	56,03
	2	22,75	5,79	18,30
IRP	0	41,62	58,44	26,53
	1	36,90	34,38	58,16
	2	21,48	7,18	15,31

4.2.1 Análise das principais variáveis explicativas II

Para efeito de análise das variáveis explicativas, foram considerados todos os dados disponíveis, pois dessa forma pode-se avaliar o potencial explicativo médio de cada variável e não somente as de um determinado ano e para tanto foi criada a Tabela 13.

Tabela 13 - Variação das probabilidades encontradas na variável classificadora ao se variar o grau de crença das variáveis explicativas segundo seus estados possíveis mantendo-se inalterados os graus de crença das demais variáveis explicativas

Grau de impacto	Variável Explicativa	Nº de estados	Baixo (%)	Mediano (%)	Superior (%)
1º	ClassAno	3	5,08	6,37-	49,70
2º	ICP	3	5,02	16,50	17,30
3º	ETPL	3	4,93	16,20	18,90
4º	IRP	3	5,42	16,20	14,10
5º	CE	3	4,51	16,10	17,90
6º	MB	3	5,40	15,80	18,50
7º	GA	2	7,26	---	17,30

A variável ICP que representa a Imobilização do capital próprio e aparece neste modelo como o segundo melhor potencial explicativo identifica a participação do Patrimônio Líquido em relação ao volume total investido no Ativo Permanente. Tal fato indica que empresas que possuem Ativos Permanentes significativos tem maiores chances de obterem rentabilidade superior . Empresas como a Petrobras, Vale do Rio doce, mineradoras em geral e empresas com grandes plantas industriais se enquadram nesse perfil. A outra variável que aparece como novidade no segundo modelo é o IRP que é o índice que identifica a participação dos recursos não correntes em relação ao volume total investido no Ativo Permanente. Esse apresenta uma característica distinta do primeiro, pois ao se variar o estado de 1 para o estado 2 a probabilidade da empresa obter rentabilidade superior decai (de 16,20% para 14,10 %) como mostrado na tabela 13. Essa característica indica que existe um limite para essa relação.

4.3 VALIDAÇÃO DO MODELO COM REGRESSÃO *LOGIT*

Seguindo os mesmos passos de Sun e Shenoy (2007) que validaram seu modelo como uma regressão *logit*, o mesmo foi feito para o presente trabalho utilizando-se como comparação o modelo 1.

Foram considerados todos os dados disponíveis (1996 até 2007) para a determinação das variáveis explicativas utilizando-se como variáveis iniciais as mesmas do modelo 2 por serem mais abrangentes ou seja, ICP, MB, CE, ETPL, IRP, CassAno e GA. Nessa regressão, as variáveis que tiveram poder explicativo foram ClassAno (Wald = 529,86 e sig.= 0,000); ICP (Wald =7,770 e sig = 0,021); e CE (Wald = 5,96 e sig. = 0,51). Tal resultado indica que as empresas que tiveram classificação superior na variável ClassAno tem 10,276 mais chances de figurarem com mais rentáveis no ano seguinte do que as que foram classificadas com rentabilidade baixa. Empresas que tiveram seu índice ICP classificado como superior e médio respectivamente tem 1,86 e 1,531 vezes mais chances de figurarem como rentabilidade superior do que as que tiveram esse índice classificado como baixo. Empresas que tiveram seu índice CE classificado como superior e médio tem respectivamente 1,490 e 1,397 vezes mais chances de figurarem como rentabilidade superior no próximo ano do que as que tiveram esse índice classificado como baixo.

Para o conjunto de empresas, no período todo, a taxa de sucesso global é mostrada na Tabela 14 e a taxas comparadas ano a ano na Tabela 15.

Tabela 14 - Classificação para o modelo *logit* utilizando todos os dados

Ocorreu efetivamente	Previsto		Total que efetivamente ocorreu	Taxa de sucesso na classificação
	RA baixa	RA superior		
RA baixa	3423	174	3597	95,16
RA superior	433	203	636	31,92
Total Previsto	3856	377	4233	85,70

Pode-se concluir que a taxa de sucesso global do modelo ao se classificar as empresas como de RA superior foi de 31,92% e ao classificar como RA baixa de 95,16% . O resultado ano a ano é mostrado na Tabela 15.

Tabela 15 - Comparativo das taxas de sucesso dos modelo 1 (RB) e a regressão *logit*

Ano	Taxas de sucesso			
	Modelo 1		Modelo *Logit*	
	Grupo 0	Grupo 1	Grupo 0	Grupo 1
1997	93,71	52,94	96,22	37,25
1998	92,41	33,80	97,83	15,49
1999	96,16	14,70	97,26	25,00
2000	95,14	31,14	97,71	14,75
2001	94,46	30,90	96,50	40,00
2002	93,37	28,07	100	5,26
2003	98,72	16,66	94,26	61,11
2004	92,60	43,39	93,56	58,49
2005	92,20	52,72	95,45	56,36
2006	93,13	39,39	96,07	24,24
Media	94,19	34,37	96,48	33,78
Desv. Padrão	2,02	13,12	1,85	19,99

Os resultados da comparação entre as médias do modelo 1 e da regressão logística sugerem que não há diferença significativa ao nível de 5%.para a previsãode RA superior, entretanto a média da previsão do modelo Logit ao prever empresas com RA baixa é superior a do modelo bayesiano (significancia de 0,037). O modelo *logit* ,na previsao da RA superior, apresenta uma dispersão maior em suas previsões (desvio padrão de 19,99 contra 13,12 do modelo bayesiano) invertendo-se tal fato na previsão da RA baixa.

Concluí-se portanto que o modelo bayesiano é tão bom quanto a regressão *Logit* para a previsão de rentabilidade superior mas tem pior desempenho ao se prever uma rentabilidade inferior.

4.4 COMPARAÇÃO ENTRE OS DOIS MODELOS BAYESIANOS

O software Netica dispõe de algumas ferramentas de análise da qualidade do modelo tais como a matriz de confusão e a qualidade do teste. Na matriz de confusão as linhas representam os dois estados que efetivamente ocorreram e as colunas, as previsões que foram feitas para estes estados. Como exemplo, a matriz de confusão da Tabela 16 representa as previsões para o ano de 2007 feitas como base nos dados de 2006, na primeira linha, e primeira coluna, o numero 287 é a quantidade de empresas que a RB previu para figurarem no estado de rentabilidade

baixa e efetivamente assim se classificaram. O numero 39 que figura na segunda linha e primeira coluna é o numero de empresa que foi previsto para estar no estado de rentabilidade superior, e que obtiveram rentabilidades superiores. Assim, na primeira linha e segunda coluna, 19 foram as empresas que foram previstas para estarem no estado rentabilidade superior e que, obtiveram rentabilidade baixa e 27 é o numero de empresas que foram corretamente previstas para o estado de rentabilidade superior. Com essa tabela é possível extrair várias informações tais como a taxa de sucesso global da previsão que é a soma dos elementos da diagonal dividido pelo total geral (314/372= 84,41%). A taxa de sucesso ao se prever que a empresa teve alta rentabilidade e isso realmente ocorreu é Tx(2)=27/66 = 40,91%. A taxa de acerto da previsão para o estado de alta lucratividade = Tx,AC=27/46=58,70% e assim sucessivamente.

A Tabela 17 apresenta a taxa de sucesso do modelo 1 a Tabela 18, a do modelo 2 e a Tabela 19 a comparação entre os dois modelos.

Tabela 16 - Matriz de confusão do primeiro modelo para o ano de 2006. Decisão com corte em p= 50%

Ocorreu efetivamente	Previsto		Total que efetivamente ocorreu
	RA baixa	RA superior	
RA baixa	287	19	306
RA superior	39	27	66
Total Previsto	326	46	372

A taxa de sucesso do grupo 2 é notoriamente inferior ao grupo 1 o que já era esperado ,uma vez que a previsão de ocorrência de uma estado que somente 18,5 % das empresas figuram seria bem mais difícil do que prever do um estado onde 81,5% das empresas se situam.

Tabela 17 - Taxa de sucesso do primeiro modelo ao longo do período para os diferentes grupos

Ano	Taxa de sucesso por grupo	
	Grupo 0 (%)	Grupo 1 (%)
1997	93,71	52,94
1998	92,41	33,80
1999	96,16	14,70
2000	95,14	31,14
2001	94,46	30,90
2002	93,37	28,07
2003	98,72	16,66
2004	92,60	43,39
2005	92,20	52,72
2006	93,13	39,39

Tabela 18 - Taxa de sucesso do segundo modelo ao longo do período para os diferentes grupos

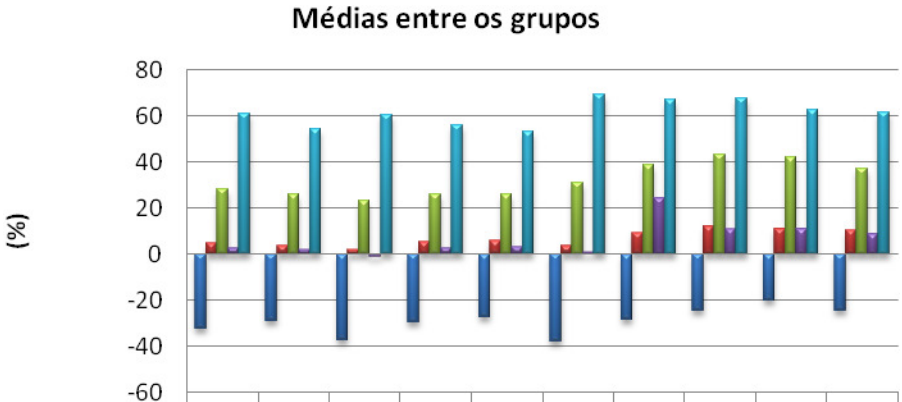
Ano	Taxa de sucesso por grupo		
	Grupo 0	Grupo 1	Grupo 2
1997	53,91	81,77	39,21
1998	63,35	77,73	25,35
1999	66,66	74,69	11,76
2000	65,28	78,60	29,50
2001	59,71	77,94	16,36
2002	63,24	72,55	28,07
2003	67,44	89,72	24,07
2004	74,31	77,22	56,60
2005	66,34	78,43	50,90
2006	69,76	77,72	42,42

Tabela 19 - Comparativa entre os dois modelos bayesianos ao se avaliar a taxa de sucesso em prever Rentabilidade Superior

Ano	Modelo 1	Modelo 2
1997	52,94	39,21
1998	33,80	25,35
1999	14,70	11,76
2000	31,14	29,50
2001	30,90	16,36
2002	28,07	28,07
2003	16,66	24,07
2004	43,39	56,60
2005	52,72	50,90
2006	39,39	42,42
Media	34,37	32,42
Desvio Padrão	13,12	14,54

De posse dos dois modelos, qual a utilidade desses para um investidor e qual o melhor? Para responder a essa pergunta é necessário a construção da utilidade para esse investidor, ou seja, quais são as expectativas de ganho considerando as probabilidades de sucesso do uso do classificador?

Como se pode ver no Gráfico 2 as diferenças entre as médias das rentabilidades (no caso do modeo 2, entre RA baixa e o de RA superior) é significativa, girando em torno de 60%, portanto a penalidade pelo fraçasso de uma aplicação em uma empresa considerando somente a perspectiva de uma retorno médio é significativa.



	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
RA baixa	-32.3	-28.8	-37.1	-29.8	-27.6	-38.2	-28.4	-24.8	-20.5	-24.4
RA média	4.99	3.82	1.95	5.26	5.91	3.5	9.23	11.93	10.92	10.29
RA superior	28.32	25.62	23.08	25.87	25.71	31.12	38.64	42.84	41.81	37.16
Média	2.53	1.91	-1.33	2.74	3.37	0.92	24.32	10.74	11.04	8.94
Diferenças	60.66	54.41	60.22	55.64	53.31	69.31	67.07	67.6	62.26	61.56

Gráfico 2- Comparativo entre as médias dos grupos e a rentabilidade média

Assim, assumindo que:

1- A expectativa de retorno de um investimento realizado aleatoriamente é a média de retorno das empresas (RA média);

2- O retorno de uma aplicação realizada em uma empresa de baixa lucratividade é a lucratividade média dos retornos das empresas de baixa lucratividade(RA baixa);

3- O retorno de uma aplicação realizada em uma empresa de alta lucratividade é a média dos retornos das empresas de alta rentabilidade (RA superior);

4- A utilidade de um investimento realizado em uma empresa é o produto da taxa de acerto da RB para esse grupo multiplicado pela RA média desse grupo.

Consequentemente, define-se as taxas de acerto como sendo a quantidade de empresas que efetivamente figuraram com a RA correspondente a seu grupo dividido pelo numero de empresas que foram previstas para este estado ou seja:

$$Taxadeacertogrupo1 = \frac{\text{nº de empresas que foram corretamente classificadas no estado1}}{\text{nº total deempresas que foram previstas para o estado1}}$$

$$Taxadeacertogrupo2 = \frac{\text{nº de empresas que foram corretamente classificadas no estado2}}{\text{nº total deempresas que foram previstas para o estado2}}$$

$$Taxadeacertogrupo3 = \frac{\text{nº de empresas que foram corretamente classificadas no estado3}}{\text{nº total deempresas que foram previstas para o estado3}}$$

A utilidade foi definida como:

$$U = U_1 + U_2 + U_3$$

onde:

$U_1 = RAmédiagrupo1 * taxadeacertogrupo1$

$U_2 = RAmédiagrupo2 * taxadeacertogrupo2$

$U_3 = RAmédiogrupo3 * taxadeacertogrupo3$

O grupo 1 é o agrupamento de empresas que obtiveram rentabilidade baixa e o grupo 2 é o agrupamento de empresas que obtiveram rentabilidade alta para o modelo 1 ou media para o modelo 2 O grupo 3 são as empresas classificadas com rentabilidade superior para o modelo 2.

A Tabela 20 apresenta a comparação entre os modelos e os cálculos das utilidades estão no Anexo C.

Tabela 20 - Comparação da utilidade para os dois modelos

Ano	U primeiro modelo %	U segundo modelo %	RA média emp. %
1997	13,03	11,90	1,91
1998	6,94	8,07	-1,33
1999	9,04	6,27	2,74
2000	12,69	14,55	3,37
2001	11,52	7,89	0,92
2002	16,71	20,83	7,64
2003	30,84	34,77	10,75
2004	22,96	28,83	10,76
2005	21,37	25,84	8,95
2006	21,08	21,66	9,89
RA acumulada	357,07	410,10	70,37

Trata-se de uma hipótes para simplesmente avaliar a diferença entre os dois modelos. O grau de risco que cada investidor está disposto a assumir é uma decisão individual e, pelo ponto de vista do problema hipotético, as RA esperadas diferem bastante, de 357,07% frente a 410,10% para o primeiro e segundo modelos respectivamente e bastante superiores a expectativa média de 70,37%. Não se pode afirmar que os dois modelos tenham taxas de sucesso distintas (sig. de 0.05) apesar do segundo apresentar um desvio padrão maior. Assim, ao se adotar as premissas das suposições iniciais o modelo 2 é superior ao primeiro pois é capaz de promover uma utilidade de rentabilidade acumulada no período de 410,10%.

5 ANÁLISES DOS RESULTADOS

O primeiro modelo apresentou uma taxa média de sucesso superior aos modelos 2 e a regressão logística, entretanto não se pode afirmar ao nível de significância de 0,05 que as taxas sejam diferentes. Mesmo assim, o primeiro modelo é mais estável devido ao seu desvio padrão menor. Os dois modelos bayesianos têm correlação de 0.805 sendo significantes ao nível de 0,01.

Uma taxa de sucesso de 14,7% no ano de 1999 revela que o modelo ainda não está estável mesmo sabendo, que nesse ano, o país passou por uma forte crise cambial. O segundo modelo também apresentou no ano de 2001 uma taxa baixa de sucesso para as previsões feitas em 2001 para os resultados de 2002 (ano da eleição do Presidente Lula). Tais fatos sugeriram que variáveis macroeconômicas seriam fundamentais para que os modelos obtenham um grau maior de estabilidade e assim foi tentada a inclusão destas.

Segundo Zurita (2008) as variáveis candidatas seriam, a taxa de juros, o cambio real, e o Pib .Esse ultimo por indicar que um aumento na atividade econômica do país permitiriam às empresas alargarem suas margens de lucro. O cambio, favoreceria as exportações e dificultaria as importações limitando a concorrência interna e causando o mesmo impacto da variável anterior sobre as margens de lucro. Os juros por sua vez contribuiriam para a redução dos custos financeiros e a facilidade para o crédito ao consumidor final.

A variável taxa de juro adotada foi a SELIC uma vez que é uma taxa referencial básica de operação no mercado financeiro, não é a realidade do mercado, mas sua variação representa um maior ou menor custo do dinheiro para as empresas e sua variação seja ascendente ou descendente é que contribuem para o aumento futuro dos custos de captação de recursos.

O calculo do cambio real foge do escopo do presente trabalho e a idéia foi utilizar uma variável *Proxi* que representasse a condição do câmbio, favorável ou não ás empresas. Foram testadas os indicadores que dizem respeito a balança comercial, tais como as exportações, as importações e o saldo da balança comercial.

A Rentabilidade média apresentou correlações significativas com as variáveis: Saldo (BC), Exportação (Exp), Importação(Imo) e juro anual (JuroAnual) no valor de 0,890, 0,840, 0,645 e -0,685 respectivamente e as correlações parciais revelaram que somente a variável Saldo e Juros deveriam ser mantidas como explicativas da Rentabilidade média (Tabela 21).

Tabela 21 - Correlação parcial entre as variáveis macroeconômicas

Variável	Variável de controlo	Correlação Parcial	Significancia
Saldo	JuroAnual	0,843	0,001
JuroAnual	Saldo	-0,514	0,087

Ao se comparar o Saldo da Balança Comercial e a Taxa de juros com as Taxas de Sucesso dos modelos, obteve-se uma correlação de 0,632 significativa ao nível de 0,05 entre o segundo modelo e o saldo da Balança Comercial. Tal resultado é um indício de que existe a possibilidade da inclusão de variáveis macroeconômicas nos modelos mesmo constatando que tal correlação somente se apresentou para o segundo e não para o primeiro modelo. Entretanto isso não foi possível alcançar no presente estudo

6 CONCLUSÕES

Foram construídos dois modelos bayesianos de previsão da lucratividade futura das empresas com base em indicadores econômico-financeiros disponíveis através dos balanços das mesmas.. A seleção dos indicadores foi baseada em uma metodologia empregada na determinação dos índices relevantes para a previsão de falência desenvolvidos por Sun e Shenoby (2007) tendo com base técnicas estatística, que avaliam a correlação entre os índices candidatos e a variável classificadora e não somente opiniões de *experts*. Tal metodologia pode ser aplicada para outros tipos de problemas de classificação ou mesmo de seleção de indicadores internos das empresas. A aplicação de redes bayesianas para esse tipo de problema se revelou intuitiva e fácil podendo ser estendido para outros tipos de problema sem grandes dificuldades.

A construção das utilidades para os modelos indicam que o emprego da ferramenta de previsão, mesmo com taxas de sucesso modestas em alguns casos pode significar uma grande diferença para os investidores (410% de rentabilidade acumulada frente a 70% da média acumulada) principalmente se nenhum conhecimento prévio tiver de uma empresa candidata a receber recursos.

Conhecimentos não númericos podem ser incorporados aos modelos e certamente causariam aumenta do poder de previsão dos mesmos.

6.1 LIMITAÇÕES

Por fugir do objetivo inicial do trabalho que era a aplicação da metodologia de seleção das variáveis e a aplicação destas num modelo bayesiano e não o aperfeiçoamento do emprego dos índices econômico-financeiros das empresas, não foram feitas considerações mais profundas sobre as variáveis selecionadas, fato este que contribuiu para a restrição do poder explicativo dos modelos. Opiniões e acréscimos de conhecimento de especialistas na área contribuirão para o

aperfeiçoamento dos modelos. Um modelo que incorpore a opinião de especialistas poderia apresentar resultados mais estáveis.

Foi constatado, que no segundo modelo, novas variáveis foram incorporadas em relação ao primeiro, ao se criar um estado intermediário para a variável classificadora. Ambas dizem respeito ao nível de investimento em ativos imobilizados das empresas. Tal fato merece um estudo mais aprofundado a respeito.

Não foi possível incluir variáveis macroeconômicas ao modelo o que pode ter comprometido a estabilidade dos mesmos e inviabilizando-os temporariamente para a construção de cenários. Construções essas que alargariam bastante a empregabilidades dos modelos.

6.2 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

A incorporação de variáveis macroeconômicas aos modelos poderia eliminar a grande dispersão nas taxas de sucesso dos mesmos tornando as previsões mais confiáveis e viabilizando a construção de cenários estratégicos para as empresas. Redes Bayesianas mais complexas como as utilizadas em Abramson (1995) talvez sejam um bom caminho para a incorporação dessas novas variáveis ao modelo Bayesiano.

As variáveis explicativas selecionadas poderiam ser melhor trabalhadas ao incorporar conhecimentos de outras áreas do conhecimento tal como a contabilidade e as finanças.

REFERÊNCIAS

ABRAMSON, B. et al. Hailfinder: a Bayesian system for forecasting severe weather. **International Journal of Forecasting**, Amsterdam, v. 12, n. 1, p. 57-71, Mar. 1996.

ABRAMSON, B.; FINIZZA, A. Probabilistic forecasts from probabilistic models: a case study in the oil market. **International Journal of Forecasting**, Amsterdam, v. 11, p. 63-72. 1995.

ALFARO, E. et al. Bankruptcy forecasting: an empirical comparison of AdaBoost and neural networks. **Decision Support Systems**, Amsterdam, v. 45, n. 1, p. 110-122, Apr. 2008.

ALTMAN, E. I. Financial ratios, discriminant analysis and prediction of corporate Bankruptcy. **Journal of Finance**, New York, v. 23, n. 4, p. 589-609, 1968.

ANDERSON, R. D. et al. A Bayesian network estimation of the service-profit chain for transport service satisfaction. **Decision Sciences**, Atlanta, v. 35, n. 4, p. 665-689, 2004.

CELIK, A. E.; KARATEPE, Y. Evaluating and forecasting banking crises through neural network models: an application for Turkish banking sector. **Expert Systems With Applications**, New York, v. 33, n. 4, p. 809-815, Nov. 2007.

CHO, S., J. et al. An integrative model with subject weight based on neural network learning for bankruptcy prediction. **Expert Systems With Applications**, New York, v. 36, n. 1, p. 403-410, Jan. 2009.

HAIR, J. F. J. et al. **Análise multivariada de dados**. Porto Alegre: Bookman, 2006.

HECHERMAN, D. **A tutorial on learning with bayesian networks**. Readmond: Microsoft Research Advanced Technology Division, 1996.

JANOT, M. M. **Modelos de previsão de insolvência bancária no Brasil**: trabalhos para discussão. Brasília: Banco Central do Brasil, 2001.

JENSEN, F. V. **Bayesian networks and decision graphs**. New York: Springer, 2007. (Statistics for engineering and information science)

KOBLIHA, M. et al. Bayesian optimization algorithms for dynamic problems. **Applications of Evolutionary Computing**, Berlin, v. 3907, p. 800-804, Mar. 2006.

KUMAR, P. R.; RAVI, V. Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review. **European Journal of Operational Research**, Amsterdam, v. 180, n. 1, p. 1-28, July 2007.

LANDAU, S.; EVERITT, B. S. **A handbook of statistical analyses using spss**: boca raton. Florida: Chapman & Hall / CRC Press LLC, 2004.

LENSBERG, T. et al. Bankruptcy theory development and classification via genetic programming. **European Journal of Operational Research**, Amsterdam, v. 169, n. 2, p. 677-697, Mar. 2006.

LI, J.; AICKELIN, U. A bayesian optimization algorithm for the nurse scheduling problem. In: PROCCEDING OF 2003 CONGRESS ON EVOLUTIONARY COMPUTATION, 2003, Camberra. **Anais...** Autralia: IEEE Press, 2003. p. 2149-2156.

MARTINS, M. S. **A previsão de insolvência pelo modelo de Cox**: uma contribuição para a análise de companhias abertas brasileiras. 2003. 103 f. Dissertação (Mestrado em Administração) - Programa de Pós-Graduação em Administração, Escola da Administração, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.

MCKEE, T. E.; LENSBERG, T. Genetic programming and rough sets: A hybrid approach to bankruptcy classification. **European Journal of Operational Research**, Amsterdam, v. 138, n. 2, p. 436-451, Apr. 2002.

NWOGUGU, M. Decision-making, risk and corporate governance: a critique of methodological issues in bankruptcy/recovery prediction models. **Applied Mathematics and Computation**, New York, v. 185, n. 1, p. 178-196, Feb. 2007.

PEARL, J. **Probabilistic reasoning in intelligent systems**: networks of plausible inference. San Mateo: Morgan Kaufmann Publishers. 1988.

PELIKAN, M. **Hierarchical bayesian optimization algorithm**. Berlin: Springer-Verlang Berlin Heidelberg. 2005.

ROCHA, F. Previsão de falência bancária: um modelo de risco proporcional. **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v. 29, n. 1, p. 137-152, abr. 1999.

SARKAR, S.; SRIRAM, R. S. Bayesian model for early warning of bank failures. **Management Science**, Providence, v. 47, n. 11, p. 1457-1475, Nov. 2001.

SHUMWAY, T. Forecasting bankruptcy more accurately: a simple hazard model. **The Journal of Business**, Chicago, v. 74, n. 1, p. 101-124, 2001.

SOHN, S. Y.; KIM, H. S. Random effects logistic regression model for default prediction of technology credit guarantee fund. **European Journal of Operational Research**, Amsterdam, v. 183, n. 1, p. 472-478, Nov. 2007.

SUN, L. L.; SHENOY, P. P. Using Bayesian networks for bankruptcy prediction: Some methodological issues. **European Journal of Operational Research**, Amsterdam, v. 180, n. 2, p. 738-753, July 2007.

TSENG, F. M.; LIN, L. A quadratic interval logit model for forecasting bankruptcy. **Omega-International Journal of Management Science**, New York, v. 33, n. 1, p. 85-91, Feb. 2005.

ZURITA, F. Bankruptcy prediction for Chilean companies. **Journal Economía Chilena**, Chile, v. 11, n. 1, p. 93-116, Apr. 2008.

ANEXOS

ANEXO A - RELAÇÃO DAS EMPRESAS ANALIZADA

521 Particip
A P Part
Abnote NM
Abyara NM
Aço Altona
Açonorte
Aços Vill
Açucar Guarani
Aes Elpa
Aes Sul
Aes Tiete
Aetatis
Agconcessoes
AGF Brasil
Agra Incorp NM
Agrale
Agrenco
Albarus
Alfa Consorc
Alfa Holding
Aliperti
All Amer Lat N2
Alpargatas N1
Altere Sec
AM Inox BR
Amadeo Rossi
Amazônia Cel
Ambev
Amelco
Americel
Amil
Ampla Energ
Anhanguera N2
Anhembi Tur
Antárctica Norte
Antárctica Polar
Antartica Paulista
Aracruz N1
Aratu
Arbra
Arcelor BR
Arno
Arteb
Arthur Lange
Azevedo
B2W
Bahema
Ban Armazens
Bardella
Battistella
Baumer
Bematech NM
Bergamo
Beta
Bic Calói
Bic Monark
Biobras
Biomm
Blue Tree
BMF
Bombril
Bompreco
Bompreço Bah
Borda do Campo
Bovespa HLD
BR Ferrovias
Br Malls Par NM
Bradespar N1
Brahma
Brampac
Brascan Res NM
Brasil T Par N1
Brasil Telec N1
Brasilagro NM
Brasilit
Braskem N1
Brasmotor
Braspérola
Brokers
BS Continental
Buettner
Bunge Alimentos
Bunge Brasil
Bunge Fertilizantes
C M A Part
C N V
Cabinda
Cacique
Caconde
Caemi
Caf Brasília
Caianda
Cambuci
Caraiba Met
Casa Anglo
Casa Jose
Casan
CBC Cartucho
CBCC
CBV
Cc Des Imob NM
CCR Rodovias NM
CEB
Cedro N1
Ceg
Celesc N2
Celg
Celm
Celpa
Celpe
Celul Irani
Cemar
Cemat
Cemepe

Cemig N1
Cesp N1
CETERP
Chapecó
Chiarelli
Cia Hering NM
Cia Providencia NM
CIBRAN
Cicanorte
Cim Itaú
CIMAF
CIMEPAR
Cimetal
Cimob
Cims
Cobrasma
Coelba
Coelce
Coest
COFAP
Coldex
Comgás
Company NM
Confab N1
Const A Lind
Const Beter
Contax
Copas
Copasa NM
Copel
Copesul
Cor Ribeiro
Cosan NM
Cosan Ltd
Cosern
Cosipa
Coteminas
CPFL Energia NM
CPFL Geração
CPFL Piratin
Cr2 NM
Cremer NM
CRT
CRT Celular
Csu Cardsyst NM
Cyrela Realt NM
Czarina
D F Vasconc
D H B
Dasa NM
Datasul NM
Dijon
Dimed
Dixie Toga
Doc Imbituba
Docas
Dohler
Douat Textil
Dova
Drogasil NM
Dtcom-Direct
Dufrybras
Duratex N1
EBE
Ecil
Ecisa
Ecodiesel NM
EDN
Eldorado
Elebra
Elekeiroz
Elekeiroz
Elektro
Eletrobras N1
Eletrolux
Eletropaulo N2
Elev Sur
Eleva
Elevad Atlas
Eluma
EMAE
Embraco
Embraer NM
Embratel Par
Encorpar
Energias BR NM
Enersul
Enxuta
EPTE
Equatorial N2
Ericsson
Escelsa
Estácio Part. N2
Estrela
Eternit NM
Eucatex
Even NM
Excelsior
Ezetec NM
F Cataguazes
F Guimaraes
Fab Bangu
Fab C Renaux
Fabrini
Fator
Fech Brasil
Fer C Atlant
Fer Haga
Fer Heringer NM
Ferbasa
Ferroban
Ferro-Ligas
Ferronorte
Fert Serrana
Fertibrás
Fertisul
FERTIZA
Fibam
Fibrasil
Flum Refrig
Forja Taurus
Fosfertil
Frangosul
Fras-Le N1
Freios Varga
Frigobras

Gafisa NM
Gazola
General Shopping NM
Geodex
Ger Paranap
Geral de Concreto
Gerdau N1
Gerdau Met N1
Giannini
Glasslite
Globex
GOL N2
Gp Invest
GPC Part
Gradiente
Granoleo
Grazziotin
Grendene NM
Grucai
GTD Part
Guaira
Guararapes
Gvt Holding NM
Habitasul
Helbor
Hércules
Hopi Hari
Hoteis Othon
IdéiasNet
Ienergia
Iguacu Café
Iguatemi NM
IKPC
Inbrac
Ind Cataguas
Ind Villares
Inds Romi NM
Inepar
Inepar Tel
Inpar NM
Invest Bemge
Invest Tur
Investec
Iochp-Maxion N1
Ipiranga Dis
Ipiranga Pet
Ipiranga Ref
Itaitinga
Itausa N1
Itautec
Iven
J B Duarte
Jaraguá Fabr
Jbs NM
Jhsf Part NM
Joao Fortes
Josapar
Karsten
Kepler Weber
Klabin N1
Klabinsegall NM
Kroton Educ. N2
Kuala
La Fonte Par
La Fonte Tel
Labo
Lacta
Lam Nacional
Lamsa
Lark Maqs
Leap
Leco
LF Tel
Liasa
Light
Light S/A NM
Lightpar
Limasa
Linh Círculo
Lisamar
Litel
Livr Globo
Lix da Cunha
Localiza NM
Log-In NM
Lojas Americ
Lojas Arapuã
Lojas Hering
Lojas Renner NM
Lopes Brasil NM
Lorenz
Lupatech NM
M G Poliest
M. Diasbranco NM
Madeirit
Magnesita
Mahle Cofap
Maio Gallo
Makro
Manasa
Mangels Indl N1
Mannesmann
Marambaia
Marcopolo N2
Marfrig NM
Marisa NM
Marisol
Martrex
Master
Matel
MCOM Wireles
Mecanica Pesada
Medial Saude NM
Mel Resistência
Melhor SP
Melpaper
Mendes Jr
Meolo
Met Duque
Metal Iguacu
Metal Leve
Metalfrio NM
Metisa
Metodo Eng
Micheletto
Microtec

Millennium
Minasmáquina
Minerva NM
Minupar
Mmx Miner NM
Moddata
Mont Aranha
Montreal
Motortec
MRV-Eng. NM
Muller
Multiplan N2
Mundial
Nadir Figuei
Nakata
Natura NM
Neoenergia
Net N2
Nordon Met
Odebrecht
Oderich
Odontoprev NM
OHL Brasil NM
Olma
Olvebra
Orion
Orniex
Oxiteno
P.Acucar-CBD N1
Pains
Panatlantica
Par Al Bahia
Pará Deminas
Paraibuna
Paranapanema
Parmalat
Paul F Luz
Pdg Realt NM
Peixe
Perdigão NM
Perdigão Agro
Persico
Pet Manguinh
Petrobrás
Petrobras Dis
Petroflex
Petropar
Petroq Uniao
Petroquisa
Pettenati
Peve Prédios
Pirelli Cabos
Pirelli Pneus
Plascar Part
Polialden
Polipropileno
Polipropileno Part
Politeno
Polpar
Porto Seguro NM
Portobello
Positivo Inf NM
Profarma NM
Pronor
Propasa
Química Ger NE
Rail Sul
Randon
Randon Part N1
Rasip Agro
Real Hold
Realpar
Recrusul
Rede Energia
Redecard NM
Renar NM
Renner Herrmann
Renner Part
Rexam
Rimet
Rio GDE Ener
Riograndense
Riosulense
Ripasa
Rodobensimob NM
Rossi Resid NM
Sabesp / NM
Sadia N1
Sadia Concordia
Saint-Gobain Canal
Saint-Gobain Vid
Salgema
Sam Industr
SAMITRI
Sanepar
Sano
Sansuy
Santanense
Santista Alim
SantisTextil
Santos Bras N2
Sao Carlos NM
São Martinho NM
Saraiva Livr N2
Satipel
Sauipe
Schlosser
Schulz
Seara Alm
Seb N2
Seg Al Bahia
Seg Min Bras
Sementes Agro
Semp
Sergen
Sharp
Shoptime
SIBRA
Sid Nacional
Sid Tubarão
Sifco
SLC Agricola NM
Sola
Sole Comex
Solorrico
Sondotecnica

Souto Vidigal
Souza Cruz
Spel Empreen
Springer
Springs Global NM
SPSCS Indl
Staroup
Submarino
Sul Amer Nac
Sul America N2
Sulacap
Sultepa
Supergasbrás
Suzano
Suzano Hold
Suzano Papel N1
Suzano Petr N2
SV Engenharia
TAM N2
TAM Transp
Tarpon
Teba
Tec Blumenau
Tecel S Jose
Technos
Tecnisa NM
Tecnosolo
TecToy
Tef Data Bra
Tegma NM
Teka
Tekno
Telasa Cel
Tele Ctr Oes
Tele Lest Cl
Tele Nord Cl
Tele Nort Cl
Tele Sudeste
Teleamapá
Telebras
Telebrasília
Teleceará Cl
Telecom Brasília
Telecoms SP
Telefonica
TELEGOIÁS
Telegoiás Cl
Telemar
Telemar N L
TELEMIG
Telemig Cl
Telemig Part
Telepisa Cl
Telerj Cl
Telern Cl
Telesc Cl
Telesp
Telesp Cl
TELEST
Telest Cl
Telet
TELMA
Telpa
Telpe Cl
Tenda NM
Tenda NM
Termin Port
Terna Part N2
Tex Renaux
Texpar
Tigre
Tim Part
Tim Sul
Totvs NM
TPI-Triunfo NM
Tractebel NM
Trafo
Tran Paulist N1
Transbrasil
Transparaná
Trevisa
Trikem
Trisul NM
Trombini
Trorion
Tupy
Ultrapar N1
Unibanco Hld N1
Unipar N1
UOL N2
Usiminas N1
Usin C Pinto
V C P N1
Vacchi
Vale R Doce N1
Varig
Varig Serv
Varig Transp
Vasp
VBC Energia
Verolme
Vicunha Sid
Vicunha Text
Vigor N1
Vivax
Vivo
Votec
Vulcabras
Weg NM
Wembley
Wentex
Wetzel
Wetzel Fund
Whirpool
White Martins
Wiest
Wilson Sons
Yara Brasil
Zivi

ANEXO B - MODELO 1 - DIFERENÇAS ENTRE AS MÉDIAS DOS DOIS GRUPOS

Teste de Levene para agualdade das variâncias				Teste t para a igualdade das médias entre os grupos		
Indice		F	Sig.	t	df	Sig. (2-tailed)
Composição do Endividamento (%)	Hipo.Variância iguais	3.6412	0.0564	-5.6846	4472.0000	0.0000
	Variâncias distintas			-5.9235	1158.6762	0.0000
Endiv. Oneroso de CP sobre Oneroso Total (%)	Hipo.Variância iguais	1.6707	0.1962	1.8777	4472.0000	0.0605
	Variâncias distintas			1.8901	1117.3307	0.0590
Endividamento Oneroso sobre P L (%)	Hipo.Variância iguais	21.0212	0.0000	-2.7932	4472.0000	0.0052
	Variâncias distintas			-1.7781	838.0541	0.0757
Endividamento Oneroso sobre Ativo Total (%)	HipoVariancia iguais	15.7042	0.0001	2.2490	4472.0000	0.0246
	Variâncias distintas			4.9341	3713.2641	0.0000
Endividamento Total sobre Ativo Total (%)	HipoVariancia iguais	1.5181	0.2180	0.6356	4472.0000	0.5251
	Variancias distintas			1.3952	3704.0035	0.1630
Endividamento Total sobre Patrimônio Líquido (%)	HipoVariancia iguais	9.9619	0.0016	-2.4236	4472.0000	0.0154
	Variancias distintas			-1.8591	911.0457	0.0633
Giro do Ativo (x)	HipoVariancia iguais	30.6416	0.0000	-6.3099	4472.0000	0.0000
	Variancias distintas			-5.2895	961.0790	0.0000
Grau de Alavancagem Financeira (x)	HipoVariancia iguais	0.8304	0.3622	0.4556	4472.0000	0.6487
	Variancias distintas			1.0000	3704.0000	0.3174
Imobilização do Capital Próprio (%)	HipoVariancia iguais	2.1205	0.1454	-1.2168	4472.0000	0.2237
	Variancias distintas			-0.9914	943.6519	0.3217

Imobilização dos Recursos Permanentes (%)	HipoVariancia iguais	1.8857	0.1698	0.1318	4472.0000	0.8952
	Variancias distintas			0.2646	4471.9412	0.7913
Liquidez Corrente (x)	HipoVariancia iguais	0.1510	0.6976	0.1748	4472.0000	0.8612
	Variancias distintas			0.1850	1178.9167	0.8533
Liquidez Geral (x)	HipoVariancia iguais	1.7858	0.1815	0.6868	4472.0000	0.4923
	Variancias distintas			1.5052	3728.4236	0.1324
Liquidez Imediata (x)	HipoVariancia iguais	3.9060	0.0482	1.0280	4472.0000	0.3040
	Variancias distintas			2.2467	3771.5525	0.0247
Liquidez Seca (x)	HipoVariancia iguais	0.1571	0.6918	0.1847	4472.0000	0.8534
	Variancias distintas			0.1955	1178.9065	0.8450
Margem Bruta (%)	HipoVariancia iguais	6.5149	0.0107	-2.6822	4472.0000	0.0073
	Variancias distintas			-3.3584	1491.0360	0.0008
Margem EBIT (%)	HipoVariancia iguais	1.3402	0.2471	0.2329	4472.0000	0.8158
	Variancias distintas			0.5028	3950.0916	0.6152
Margem EBITDA (%)	HipoVariancia iguais	1.4592	0.2271	-0.6190	4472.0000	0.5360
	Variancias distintas			-1.3569	3724.8856	0.1749
Margem Líquida (%)	HipoVariancia iguais	3.1683	0.0751	-0.9333	4472.0000	0.3507
	Variancias distintas			-2.0102	3979.2971	0.0445
Margem Operacional (%)	HipoVariancia iguais	3.1083	0.0780	-0.9392	4472.0000	0.3477
	Variancias distintas			-2.0217	3986.6071	0.0433
Rentabilidade do Ativo Total (%)	HipoVariancia iguais	2.2468	0.1340	-0.7694	4472.0000	0.4417
	Variancias distintas			-1.6889	3704.0163	0.0913
Retorno das Participações (%)	HipoVariancia iguais	1.0955	0.2953	-0.4337	4472.0000	0.6646
	Variancias distintas			-0.9517	3708.6058	0.3413
Retorno do Acionista (%)	HipoVariancia iguais	22.5908	0.0000	-2.3755	4472.0000	0.0176
	Variancias distintas			-1.0818	768.0000	0.2797
Retorno do Investimento Total (%)	HipoVariancia iguais	2.2362	0.1349	-0.7675	4472.0000	0.4428

	Variancias distintas			-1.6849	3704.0157	0.0921
Termômetro Financeiro (x)	HipoVariancia iguais	0.7288	0.3933	-0.4045	4472.0000	0.6859
	Variâncias distintas			-0.8668	4044.7882	0.3861

ANEXO C - SEGUNDO MODELO: DIFERENÇAS ENTRE AS MÉDIAS DOS 3 GRUPOS TESTE LSD. IRP E MB NÃO POSSUEM MEDIAS SIGNIFICATIVAMENTE DIFERENTES ENTRE DOIS DE SEUS GRUPOS, MAS FORAM MANTIDAS NO MODELO

Variável	Grupo A	Grupo B	Diferença	Significância
CE	0	1	-5,81	0,000
		2	-9,35	0,000
	1	0	5,81	0,000
		2	-3,54	0,005
	2	0	9,35	0,000
		1	3,54	0,005
ETPL	0	1	116,47	0,000
		2	34,60	0,169
	1	0	-116,47	0,000
		2	81,86	0,000
	2	0	25,18	0,169
		1	81,86	0,000
GA	0	1	-0,05751	0,014
		2	-0,2289	0,000
	1	0	0,0575	0,014
		2	-0,1714	0,000
	2	0	0,2289	0,000
		1	0,1714	0,000
ICP	0	1	47,9984	0,000
		2	23,4459	0,010
	1	0	-47,9984	0,000
		2	-24,5524	0,003
	2	0	-234459	0,010
		1	-245524	0,003
IRP	0	1	13,9730	0,000
		2	10,3303	0,039
	1	0	-13,9730	0,000
		2	-3,6427	0,422
	2	0	-10,3303	0,039
		1	3,3427	0,422
MB	0	1	-1,2023	0,210
		2	-4,0918	0,002
	1	0	1,2023	0,210
		2	-2,8895	0,016
	2	0	4,0918	0,002
		1	2,8895	0,016

ANEXO D - RESUMO DO MODELO 1

	1997					**1998**					**1999**			
Real	Prev 0	Prev 1	Total	Tx sucesso	Real	Prev 0	Prev 1	Total	Tx sucesso	Real	Prev 0	Prev 1	Total	**Tx sucesso**
Prev 0	298.00	20.00	318.00	0.94	Prev 0	341.00	28.00	369.00	0.92	Prev 0	351.00	14.00	365.00	**0.96**
Prev. 1	24.00	27.00	51.00	0.53	Prev. 1	47.00	24.00	71.00	0.34	Prev. 1	58.00	10.00	68.00	**0.15**
	Total	47.00				Total	52.00				Total	24.00		
	Tx acerto2	0.57				Tx acerto2	0.46				Tx acerto2	0.42		
	Tx acerto2	0.43				Tx acerto2	0.54				Tx acerto2	0.58		
	2000					**2001**					**2002**			
Real	Prev 0	Prev 1	Total	Tx sucesso	Real	Prev 0	Prev 1	Total	Tx sucesso	Real	Prev 0	Prev 1	Total	**Tx sucesso**
Prev 0	333.00	17.00	350.00	0.95	Prev 0	324.00	19.00	343.00	0.94	Prev 0	310.00	22.00	332.00	**0.93**
Prev. 1	42.00	19.00	61.00	0.31	Prev. 1	38.00	17.00	55.00	0.31	Prev. 1	41.00	16.00	57.00	**0.28**
	Total	36.00					36.00				Total	38.00		
	Tx acerto2	0.53				Tx acerto2	0.47				Tx acerto2	0.42		
	Tx acerto2	0.47				Tx acerto2	0.53				Tx acerto2	0.58		
	2003					**2004**					**2005**			
Real	Prev 0	Prev 1	Total	Tx sucesso	Real	Prev 0	Prev 1	Total	Tx sucesso	Real	Prev 0	Prev 1	Total	Tx sucesso
Prev 0	310.00	4.00	314.00	0.99	Prev 0	288.00	23.00	311.00	0.93	Prev 0	284.00	24.00	308.00	0.92
Prev. 1	45.00	9.00	54.00	0.17	Prev. 1	30.00	23.00	53.00	0.43	Prev. 1	26.00	29.00	55.00	0.53
	Total	13.00				Total	46.00				Total	53.00		
	Tx acerto2	0.69				Tx acerto2	0.50				Tx acerto2	0.55		
	Tx acerto2	0.31				Tx acerto2	0.50				Tx acerto2	0.45		

	2006			
	Prev 0	Prev 1	Total	Tx sucesso
Prev 0	285.00	21.00	306.00	0.93
Prev. 1	40.00	26.00	66.00	0.39
	Total	47.00		
	Tx acerto2	0.55		
	Tx acerto2	0.45		

ANEXO E – RESULTADOS DO MODELO LOGIT

	Ano 1996			
	Previsto		Total	Perc
Observado	0	1		
0	271	10	281	0,9644
1	23	27	50	0,54
Total	294	37	331	
Perc	0,0782	0,7297		

	Ano 1997			
	Previsto		Total	Perc
Observado	0	1		
0	306	12	318	0,9623
1	32	19	51	0,3725
Total	338	31	369	
Perc	0,0947	0,6129		

	Ano 1998			
	Previsto		Total	Perc
Observado	0	1		
0	361	8	369	0,9783
1	60	11	71	0,1549
Total	421	19	440	
Perc	0,1425	0,5789		

	Ano 1999			
	Previsto		Total	Perc
Observado	0	1		
0	355	10	365	0,9726
1	51	17	68	0,25
Total	406	27	433	
Perc	0,1256	0,6296		

	Ano 2000			
	Previsto		Total	Perc
Observado	0	1		
0	342	8	350	0,9771
1	52	9	61	0,1475
Total	394	17	411	
Perc	0,132	0,5294		

	Ano 2001			
	Previsto		Total	Perc
Observado	0	1		
0	331	12	343	0,965
1	33	22	55	0,4
Total	364	34	398	
Perc	0,0907	0,6471		

	Ano 2002			
	Previsto		Total	Perc
Observado	0	1		
0	332	0	332	1
1	54	3	57	0,0526
Total	386	3	389	
Perc	0,1399	1		

	Ano 2003			
	Previsto		Total	Perc
Observado	0	1		
0	296	18	314	0,9427
1	21	33	54	0,6111
Total	317	51	368	
Perc	0,0662	0,6471		

	Ano 2004			
	Previsto		Total	Perc
Observado	0	1		
0	291	20	311	0,9357
1	22	31	53	0,5849
Total	313	51	364	
Perc	0,0703	0,6078		

	Ano 2005			
	Previsto		Total	Perc
Observado	0	1		
0	294	14	308	0,9545
1	24	31	55	0,5636
Total	318	45	363	
Perc	0,0755	0,6889		

	Ano 2006			
	Previsto		Total	Perc
Observado	0	1		
0	291	15	306	0,951
1	39	27	66	0,4091
Total	330	42	372	
Perc	0,1182	0,6429		

ANEXO F - RESUMO MODELO 2

Ano					Tx sucesso						Tx sucesso
1997	62	49	4	115	0,5391	1998	83	45	3	131	0,6335878
	22	166	15	203	0,8177		31	185	22	238	0,7773109
	5	26	20	51	0,3922		10	43	18	71	0,2535211
Total	89	241	39	369		Total	124	273	43	440	
					Tx sucesso						Tx sucesso
1999	80	36	4	120	0,6667	2000	79	41	1	121	0,6528926
	44	183	18	245	0,7469		29	180	20	229	0,7860262
	11	49	8	68	0,1176		7	36	18	61	0,295082
Total	135	268	30	433		Total	115	257	39	411	
					Tx sucesso						Tx sucesso
2001	83	53	3	139	0,5971	2002	74	42	1	117	0,6324786
	31	159	14	204	0,7794		37	156	22	215	0,7255814
	3	43	9	55	0,1636		16	25	16	57	0,2807018
Total	117	255	26	398		Total	127	223	39	389	
					Tx sucesso						Tx sucesso
2003	87	41	1	129	0,6744	2004	81	27	1	109	0,7431193
	17	166	2	185	0,8973		27	156	19	202	0,7722772
	3	38	13	54	0,2407		7	16	30	53	0,5660377
Total	107	245	16	368		Total	115	199	50	364	
					Tx sucesso						Tx sucesso
2005	69	33	2	104	0,6635	2006	60	21	5	86	0,6976744
	29	160	15	204	0,7843		35	171	14	220	0,7772727
	5	22	28	55	0,5091		9	29	28	66	0,4242424
Total	103	215	45	363		Total	104	221	47	372	

ANEXO G - UTILIDADE MODELO 1

Ano	Tx 1	(1-tx1)	Ra (0)	Ra (1)	U1	U2	Ut		MediaRA		Prevemp grupo 1	Acertos grupo2
1997	0,57	0,43	-3,14	25,58	-1,3502	14,581	13,23	1,1323	1,91	1,0191	47	27
1998	0,46	0,54	-3,61	23,08	-3,672	10,617	6,9448	1,2109	-1,33	1,0055	52	24
1999	0,41	0,59	-6,8	25,87	-1,5635	10,607	9,0432	1,3204	2,74	1,0331	24	10
2000	0,52	0,48	-2,65	25,71	-0,6768	13,369	12,692	1,488	3,37	1,0679	36	19
2001	0,47	0,53	-1,41	31,12	-3,0952	14,626	11,531	1,6596	0,92	1,0777	36	17
2002	0,42	0,58	-5,84	38,64	0,4756	16,229	16,704	1,9369	7,64	1,1601	38	16
2003	0,69	0,31	0,82	43,11	1,0912	29,746	30,837	2,5341	10,75	1,2848	13	9
2004	0,5	0,5	3,52	41,81	2,055	20,905	22,96	3,116	10,76	1,423	46	23
2005	0,54	0,46	4,11	37,16	1,3018	20,066	21,368	3,7818	8,95	1,5504	53	29
2006	0,55	0,45	2,83	34,37	0	18,904	18,904	4,4967	9,89	1,7037	47	26
							total	349,67		70,372	392	200

ANEXO H - UTILIDADE MODELO 2

Ano	Tx1	Tx2	Tx3	Ra (0)	Ra (1)	RA (2)	U1	U2	U3	Ut	MediaRA	NºPreV.	NºAcertos
1997,00	0,10	0,38	0,52	-28,79	3,83	25,62	-2,88	1,46	13,32	11,90	1,91	39,00	20,00
1998,00	0,07	0,51	0,42	-37,14	1,92	23,08	-2,60	0,98	9,69	8,07	-1,33	43,00	18,00
1999,00	0,13	0,60	0,27	-29,77	5,26	25,87	-3,87	3,16	6,98	6,27	2,74	30,00	8,00
2000,00	0,02	0,51	0,47	-27,61	5,91	25,71	-0,55	3,01	12,08	14,55	3,37	39,00	18,00
2001,00	0,12	0,54	0,34	-38,19	3,50	31,12	-4,58	1,89	10,58	7,89	0,92	26,00	9,00
2002,00	0,02	0,56	0,42	-28,44	9,23	38,65	-0,57	5,17	16,23	20,83	7,64	39,00	16,00
2003,00	0,06	0,13	0,81	-24,76	11,93	42,85	-1,49	1,55	34,71	34,77	10,75	16,00	13,00
2004,00	0,02	0,38	0,60	-20,45	10,93	41,81	-0,41	4,15	25,09	28,83	10,76	50,00	30,00
2005,00	0,04	0,33	0,63	-24,41	10,29	37,17	-0,98	3,40	23,42	25,84	8,95	45,00	28,00
2006,00	0,11	0,29	0,60	-16,49	9,82	34,37	-1,81	2,85	20,62	21,66	9,89	47,00	28,00
								Total acumulado		410,10		374,00	188,00