



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, ATUÁRIA E**  
**CONTABILIDADE**  
**DEPARTAMENTO DE ADMINISTRAÇÃO**  
**CURSO DE CIÊNCIAS ATUARIAIS**

**FABRICIO FREITAS DO NASCIMENTO**

**DETERMINAÇÃO DOS MELHORES INDICADORES CONTÁBEIS E TÉCNICAS**  
**DE ANÁLISE MULTIVARIADA PARA PREVER INSOLVÊNCIA DE**  
**OPERADORAS DE SAÚDE MÉDICO-HOSPITALARES**

**FORTALEZA**

**2017**

FABRICIO FREITAS DO NASCIMENTO

DETERMINAÇÃO DOS MELHORES INDICADORES CONTÁBEIS E TÉCNICAS DE  
ANÁLISE MULTIVARIADA PARA PREVER INSOLVÊNCIA DE OPERADORAS DE  
SAÚDE MÉDICO-HOSPITALARES

Monografia apresentada à Faculdade de  
Economia, Administração, Atuária,  
Contabilidade e Secretariado, como requisito  
parcial para obtenção do Grau de Bacharel em  
Ciências Atuariais.

Orientador(a): Prof. Ms. Rômulo Alves  
Soares.

FORTALEZA

2017

FABRICIO FREITAS DO NASCIMENTO

DETERMINAÇÃO DOS MELHORES INDICADORES CONTÁBEIS E TÉCNICAS DE ANÁLISE MULTIVARIADA PARA PREVER INSOLVÊNCIA DE OPERADORAS DE SAÚDE MÉDICO-HOSPITALARES

Monografia apresentada à Faculdade de Economia, Administração, Atuária, Contabilidade e Secretariado, como requisito parcial para obtenção do Grau de Bacharel em Ciências Atuariais.

Aprovada em: \_\_\_\_ / \_\_\_\_ / \_\_\_\_\_.

BANCA EXAMINADORA

---

Prof. Ms. Rômulo Alves Soares (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Ms. Lineker Costa Passos  
Instituto Federal do Piauí (IFPI)

---

Prof. Ms. Lahis Muriel Feliciano dos Santos  
Faculdade Lourenço Filho (FLF)

A todos que, de alguma forma, contribuem para o meu crescimento pessoal, acadêmico e profissional.

## AGRADECIMENTOS

Agradeço a Deus pelo dom da vida. À minha mãe que sempre se mostrou um exemplo de ser humano e é muito fiel aos seus propósitos, além de encarar todos os problemas com muita paciência na certeza de que tudo dará certo. Ao meu pai por simplesmente existir e me motivar com a sua existência a batalhar por todas as nossas conquistas enquanto família. Ao meu irmão, Fábio, pela sua simplicidade e boa forma de ver as coisas ao seu redor. À minha avó, Francisca por cuidar tão bem de todos ao seu redor. À minha prima Keyla, por todo incentivo, companheirismo e quebrações de galhos.

Agradeço ao meu orientador, Rômulo, pelo exemplo de cientista de dados que é e por toda a paciência e suporte na construção desse trabalho.

À todos os amigos e companheiros de trabalho da Unimed Fortaleza, sobretudo à Gerência de Informações Estratégicas e Atuariais - GIEAT, onde aprendi grande parte do que sei sobre a vida profissional, a atuária em si e sobre inspiração, ao Escritório de Planejamento, Gestão e Desenvolvimento Organizacional – EPLAN, e em específico ao Team Marcelo onde aprendo sobre disciplina, autocontrole, foco e o melhor de tudo, sem métodos tradicionais de ensino.

Aos grandes amigos que a atuária me deu, dos quais vale destacar o André, que sempre foi meu braço direito em toda a jornada acadêmica, que dividiu grandes momentos da vida comigo, que representou, e muito bem, parte da minha família nessa correria. Ao Ronery e à Vivi que comigo formam um trio que vocês ainda ouvirão muito falar. À Larissa, por ser um grande ser humano. Ao Carlim, a Viviane, ao Thales, ao Anderson e ao Vitor por serem quem são.

Agradeço também ao meu grande amigo Paulo Maia, por sempre me motivar e acreditar que eu posso superar qualquer desafio, à Neyla por ser alguém que entrou na minha vida pra morar e faz bom uso dessa estadia e ao Luiz que por onde passa deixa um pouco de alegria.

“A falha é simplesmente a oportunidade de  
começar de novo de forma mais inteligente.”

(Henry Ford)

## RESUMO

O sistema de saúde suplementar se tornou essencial na vida da sociedade moderna, uma vez que o Sistema Único de Saúde – SUS apresenta diversas deficiências que vão desde a superlotação das unidades hospitalares, até a falta de materiais e medicamentos a serem fornecidos aos pacientes. Visto a importância do serviço ofertado pelas operadoras de saúde médico-hospitalares, observar o seu desempenho com relação a solvência é primordial, uma vez que a ruína destas empresas causa prejuízos significativos a vida de seus beneficiários. Diante deste cenário, esta pesquisa busca, através de técnicas de análise multivariada, identificar quais indicadores contábeis mais se relacionam com o estado de solvência de uma OPS e qual das técnicas desempenha melhor papel na previsão de insolvência. Para o experimento, foram selecionadas 177 operadoras de saúde aleatoriamente, sendo 59 insolventes e 118 solventes e sobre os indicadores contábeis destas foram aplicados as técnicas de regressão logística, árvores de classificação e regressão, redes neurais artificiais e *random forest*. Os resultados do estudo apontam que o indicador de Liquidez Geral é o que mais se relacionam com o estado de solvência das operadoras e que as redes neurais artificiais apresentam melhor performance na previsão deste cenário.

**Palavras-chave:** Previsão de Insolvência. Operadoras de Saúde. Análise Multivariada.

## **ABSTRACT**

The supplementary health system became essential in the life of modern society. Since the public health system has several deficiencies ranging from overcrowding of the hospital units to the lack of materials and medicines to be provided to patients. Given the importance of the service offered by medical-hospital health care providers, observing their performance in relation to solvency is paramount, because the ruin of these companies cause significant losses to their beneficiaries. On this, this research seeks, through multivariate analysis techniques, to identify which accounting indicators are most relate to the solvency status of a health care provider and which of the techniques plays a better role predicting insolvency. For the experiment, 177 health care providers were randomly selected, 59 of which were insolvent and 118 were solvents. The logistic regression techniques, classification and regression trees, artificial neural networks and random forest were applied to the accounting indicators. The results of the study indicate that the indicator of General Liquidity are the one that most relate to the solvency of the operators and that the artificial neural networks present better performance in the forecast of this situation.

**Keywords:** Predicting insolvency. Health care providers. Multivariate analysis.



## LISTA DE FIGURAS

Figura 1 - Estrutura de uma MLP .....	29
---------------------------------------	----

## **LISTA DE GRÁFICOS**

Gráfico 1 - Relação de Operadoras x Beneficiários (2000-2015) .....	17
Gráfico 2 - Curvas ROC dos Modelos .....	33

## LISTA DE QUADROS

Quadro 1 – Fórmula discriminante de Altman (1968) .....	21
Quadro 2 – Resumo do modelo proposto por Guimarães e Alves (2009).....	22
Quadro 3 – Indicadores Utilizados .....	25

## LISTA DE TABELAS

Tabela 1 – Distribuição de Beneficiários por Modalidades de OPS – Mar/2017.....	19
Tabela 2 - Estatísticas descritivas dos dados .....	31
Tabela 3 - Grau de importância das variáveis por técnica.....	32
Tabela 4 - Classificação do Modelo - Amostra de Treino.....	32
Tabela 5 - Classificação do Modelo - Amostra de Teste 1.....	34
Tabela 6 - Classificação do Modelo - Amostra de Teste 2.....	34
Tabela 7 - Classificação do Modelo - Amostra de Teste 3.....	35

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>14</b>
<b>2 REFERENCIAL TEÓRICO .....</b>	<b>16</b>
<b>2.1 Histórico do Mercado de Saúde Suplementar.....</b>	<b>16</b>
<b>2.2 Classificação das Operadoras .....</b>	<b>18</b>
<b>2.3 Estudos de Previsão de Insolvência e Indicadores Contábeis.....</b>	<b>19</b>
<b>3 METODOLOGIA .....</b>	<b>24</b>
<b>3.1 Tipo de Pesquisa.....</b>	<b>24</b>
<b>3.2 População e Amostra .....</b>	<b>24</b>
<b>3.3 Coleta, Organização e Análise dos Dados.....</b>	<b>25</b>
<b>3.4.1 Regressão Logística.....</b>	<b>26</b>
<b>3.4.2 Árvores de Regressão e Classificação .....</b>	<b>27</b>
<b>3.4.3 Randon Forest.....</b>	<b>28</b>
<b>3.4.4 Redes Neurais Artificiais .....</b>	<b>29</b>
<b>3.4.4 Curvas ROC.....</b>	<b>30</b>
<b>4 RESULTADOS .....</b>	<b>31</b>
<b>4.1 Estatísticas Descritivas .....</b>	<b>31</b>
<b>4.2 Técnicas de Análise Multivariada .....</b>	<b>31</b>
<b>4.2.1 Variáveis Predictoras de Insolvência .....</b>	<b>31</b>
<b>4.2.2 Desempenho das Técnicas na Previsão de Insolvência.....</b>	<b>32</b>
<b>5. CONCLUSÃO.....</b>	<b>36</b>
<b>6. REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>38</b>
<b>APÊNDICES .....</b>	<b>38</b>
<b>APÊNDICE A – Operadoras Utilizadas no Estudo.....</b>	<b>40</b>
<b>APÊNDICE B – Códigos Utilizados no R.....</b>	<b>43</b>

## 1 INTRODUÇÃO

Frente à magnitude do valor fornecido à sociedade por um sistema de saúde alternativo ao Sistema Único de Saúde - SUS, é de fundamental importância atentar-se para a situação econômico-financeira das Operadoras de Planos de Assistência à Saúde – OPS, sobretudo no que diz respeito a sua solvência, uma vez que a falência e a consequente suspensão dos serviços de pelo menos uma dessas instituições pode causar impactos sensíveis ao mercado de saúde suplementar.

Como exemplo disso, pode ser citado o caso da Unimed Paulistana, cooperativa médica que possuía uma das maiores carteiras de clientes do sistema Unimed com mais de 744 mil beneficiários (ANS, 2014), e que desde 2014 apresentava problemas com seus credores, o que levou a instituição à liquidação extrajudicial (sansão da Agência Nacional de Saúde Suplementar - ANS que retira a OPS do mercado de saúde suplementar, impedindo-a de comercializar seus produtos) em fevereiro de 2016. Com a liquidação, os beneficiários da operadora tiveram que ser transferidos para outras empresas do ramo, o que resultou em prejuízo para estes, visto que não foi possível encontrar oferta de planos com coberturas semelhantes aos que usufruíam na Unimed Paulistana com uma tabela de preços parecida.

Além disso, o mercado de saúde suplementar é extremamente competitivo. Segundo o Caderno de Informações da Saúde Suplementar da ANS de 2016, o número de operadoras de planos de saúde médico-hospitalares com beneficiários em atividade reduziu cerca de 43% desde 2000, em contrapartida, este mesmo relatório reporta um aumento de mais de 60% na quantidade de beneficiários de assistência médica com ou sem odontologia.

Ou seja, frente a rigorosidade das exigências legais, dentre outros fatores, tem-se uma elevação na quantidade de operadoras que não conseguem se sustentar em um ambiente altamente regulado e consequentemente deixam seus beneficiários a disposição do mercado. Por outro lado, há um aumento na procura por planos de saúde nas operadoras que conseguem se sustentar. Desse modo, prover maneiras de garantir que as OPS com registro ativo se mantenham sustentáveis é essencial.

Baseado neste cenário, ferramentas capazes de antever situações de dificuldades financeiras e apontar suas razões, ganham importância como formas de aumentar as chances de continuidade das empresas. FitzPatrick (1932) foi o pioneiro nos estudos de previsão de insolvência com base na utilização de indicadores contábeis. Posteriormente, Beaver (1966) tentou prever a insolvência de empresas comerciais utilizando, além de indicadores contábeis, mecanismos de análise univariada.

Em seguida Altman (1968) desenvolveu um modelo que utilizava análise discriminante múltipla para previsão de insolvência de 66 empresas, sendo 33 solventes e 33 insolventes. Este modelo foi um marco nos estudos deste tema, visto que corrigia alguns problemas do estudo desenvolvido por Beaver. Ohlson (1980) optou por utilizar regressão logística condicional, sendo o pioneiro na utilização da técnica com o intuito de prever insolvência. A partir desses estudos, a utilização de análise multivariada em conjunto com indicadores contábeis passou a ser uma das principais formas de se prever insolvência.

No cenário nacional, os primeiros estudos focados na previsão de insolvência foram de Kantiz (1976), este identificou através de análise de balanços, quais as variáveis com maior poder de separar empresas solventes de insolventes. Elizabetsky (1976) foi o primeiro autor brasileiro a utilizar análise multivariada para classificar clientes de um banco. Algumas décadas depois, Guimarães e Alves (2009) desenvolveram um modelo capaz de prever a insolvência de operadoras de saúde nos próximos doze meses através de regressão logística.

Apesar da grande quantidade de aplicações técnicas multivariadas considerando um contexto de previsão de insolvência, esses estudos muitas vezes focam apenas na capacidade preditiva, e não no que é importante para a diferenciação de empresas solventes e insolventes.

Diante do exposto, esta pesquisa tem como objetivo verificar quais variáveis melhor separam Operadoras de Planos de Assistência à Saúde solventes das insolventes. De modo específico, busca-se aplicar técnicas de análise multivariada para a previsão de insolvência de operadoras de planos de saúde, e comparar o desempenho obtido com essas aplicações.

Para isto serão utilizadas quatro técnicas de análise multivariada: regressão logística, árvores de regressão e classificação, redes neurais artificiais e *random forest* aplicados a indicadores contábeis sugeridos por Soares (2006). Os indicadores foram calculados de acordo com os dados disponibilizados pela agência nacional de saúde suplementar através do Documento de Informações Periódicas das Operadoras de Plano de Assistência à Saúde – DIOPS dos anos de 2014 a 2016, sendo o primeiro e o segundo ano utilizados para estimar o modelo e o segundo e o terceiro utilizados para testá-lo.

## **2 REFERENCIAL TEÓRICO**

Nesta seção será apresentado um histórico do mercado de saúde suplementar no Brasil e seu desenvolvimento, os tipos de operadoras classificados pelo órgão regulador de acordo com seu estatuto jurídico. Posteriormente, serão apresentados alguns estudos envolvendo insolvência, por fim, aplicações de técnicas de previsão de insolvência sobre indicadores contábeis.

### **2.1 Histórico do Mercado de Saúde Suplementar**

Desde a década de 50, empresas do setor privado direcionam atenção especial ao ramo de saúde. O manual “Regulação e Saúde” do Ministério da Saúde (2002) relata que as instituições públicas e privadas constituíam fundos a partir de seus recursos e dos recursos de seus colaboradores destinados a cobrir eventuais despesas com assistência médica, assim, deu-se início a expansão do mercado de saúde suplementar.

Algum tempo depois, a constituição federal de 1988 garantia o direito a todos os cidadãos ao acesso a saúde, ao afirmar nos termos do art.196 que “A saúde é direito de todos e dever do estado[.]” e nos artigos posteriores que as ações e serviços de saúde são de relevância pública, mas também podem ser executados por pessoa física ou jurídica de direito privado. Desse modo, ficou reservada a iniciativa privada a atuação no setor de saúde de forma suplementar, ainda sem regulamentação específica destinada a essa fatia da seguridade social, que, por ora, estava sob responsabilidade do Ministério da Previdência Social.

Segundo Sestelo et al (2013), na década de 90, a visão estado x mercado, deu lugar a uma mescla de público/privado na prestação dos serviços de saúde, de modo que o Estado assumiria um papel de regulador do mercado e não necessariamente de único provedor de serviços. Isto se deu sobretudo devido a percepção das deficiências do Sistema Único de Saúde - SUS, criado também nesta década, as fraquezas da saúde pública foi o fator primordial para o surgimento de várias empresas que pudessem suportar juntamente com o Estado a demanda por assistência à saúde advinda da sociedade. Outro acontecimento importante da época foi a extinção do Instituto Nacional de Assistência Médica da Previdência Social – INAMPS, conseqüentemente transferindo a responsabilidade da regulação do setor para o Ministério da Saúde.

Com a expansão da atividade das Operadoras de Saúde, havia consenso que era necessário uma maior intervenção estatal sobre os serviços prestados quanto a sua regulamentação, dessa forma o marco regulatório para o setor de saúde suplementar se deu a partir da publicação da lei 9.656/98 que regulamenta os planos e seguros privados de



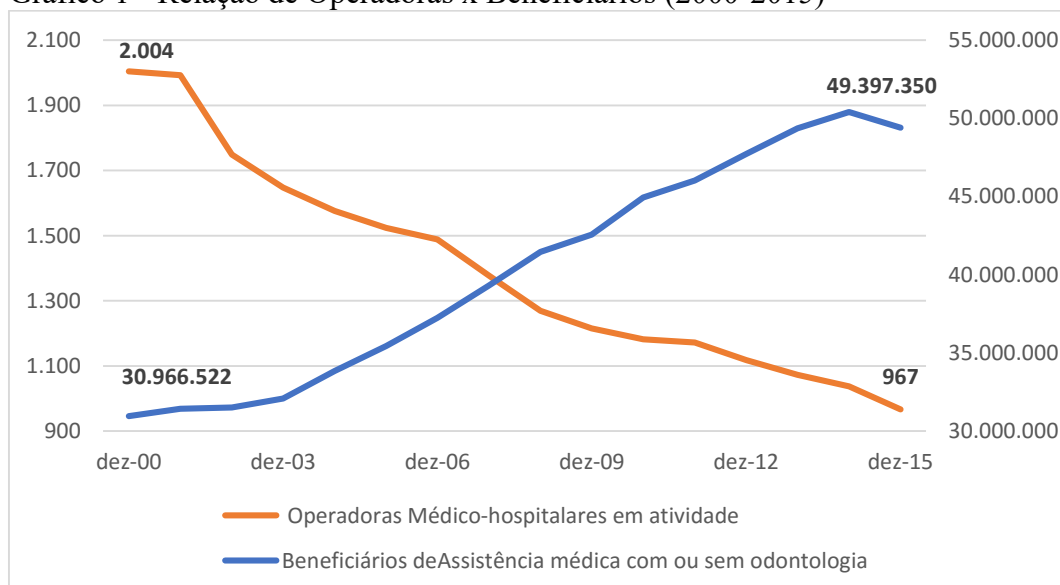
assistência à saúde, segundo o texto da própria lei. A norma dá instruções sobre as relações entre beneficiários, operadoras e prestadores de serviços de saúde, além de regulamentar os contratos de prestação de serviços de saúde entre pessoas físicas e jurídicas e dispor também sobre prazos e garantias dos agentes envolvidos neste mercado. Segundo Barros e Beiruth (2016, v.4, p.17):

Antes da publicação da Lei 9656/98, que regulamentou o setor e que precede a criação da ANS, até meados 1994, os serviços de seguros-saúde, planos de saúde e assistência médica eram livres para arbitrar sobre preços, coberturas e carências e decidiam sobre quais contratos eram interessantes para si, pois podiam rescindi-los unilateralmente.

Dois anos depois da publicação desta norma, a Agência Nacional de Saúde Suplementar foi constituída por meio da lei 9.961/00, e seu surgimento trouxe importantes transformações para o setor. O início da atuação da autarquia foi marcado pela publicação de normas do qual o assunto tema era a perspectiva econômico-financeira das operadoras. A rigorosidade das exigências legais aliada a elevação dos custos com serviços de saúde, retratada por Guimarães e Alves (2009), Leal e Matos (2009) e Salles Dias Filho (2010) em seus estudos, fez com que muitas destas operadoras não conseguissem se sustentar neste mercado.

Dados da Agência mostram que em dez/15, 15 anos após sua criação, o número de operadoras médico-hospitalares com beneficiários em atividade no Brasil reduziu cerca de 43%. Este número se torna ainda mais expressivo quando comparado com o histórico de beneficiários do mesmo período que, por sua vez, cresceu aproximadamente 60% chegando a 49,4 milhões, conforme expresso no gráfico abaixo:

Gráfico 1 - Relação de Operadoras x Beneficiários (2000-2015)



Fonte: Adaptado pelo autor com base em ANS (2016)

A relação entre operadoras e beneficiários tem se comportado em ordem inversamente proporcional nos últimos anos, enquanto as primeiras estão cada vez em menor proporção, os segundos aumentam vertiginosamente, o que eleva a competitividade de um mercado que além de ter como obrigação cuidar da saúde de seus clientes, deve se preocupar também com sua saúde financeira para se manter sustentável.

## 2.2 Classificação das Operadoras

As operadoras de saúde se diferenciam entre si de acordo com seu estatuto jurídico, assim, também atendem a regulações específicas conforme sua classificação. As OPS são distribuídas em 8 categorias principais sendo elas (ANS, 2014):

**Autogestão:** empresa que opera planos de assistência à saúde destinados, exclusivamente, a empregados ativos, aposentados, pensionistas ou ex-empregados, de uma ou mais empresas ou, ainda, a participantes e dependentes de associações de pessoas físicas ou jurídicas, fundações, sindicatos, entidades de classes profissionais ou assemelhados e seus dependentes.

**Cooperativa médica:** sociedade sem fim lucrativo, conforme o disposto na Lei nº 5.764, de 16 de dezembro de 1971.

**Cooperativa odontológica:** sociedade sem fim lucrativo, conforme o disposto na Lei nº 5.764, de 16 de dezembro de 1971, que opera exclusivamente planos odontológicos.

**Filantropia:** entidade sem fins lucrativos que opera planos privados de assistência à saúde, sendo certificada como entidade filantrópica junto ao Conselho Nacional de Assistência Social (CNAS) e declarada de utilidade pública pelo Ministério da Justiça ou pelos órgãos dos governos estaduais e municipais.

**Administradora:** empresa que apenas administra planos de saúde, que são financiados por outra operadora. Uma administradora não assume o risco decorrente da operação desses planos e não possui rede própria, credenciada ou referenciada de serviços médico-hospitalares ou odontológicos.

**Seguradora especializada em saúde:** sociedade seguradora autorizada a operar planos de saúde, desde que esteja constituída como seguradora especializada nesse tipo de seguro, devendo seu estatuto social vedar a atuação em quaisquer outros ramos ou modalidades.

**Medicina de grupo:** demais empresas ou entidades que operam planos privados de assistência à saúde.

**Odontologia de grupo:** demais empresas ou entidades que operam, exclusivamente, planos odontológicos.

Conforme dados de março de 2017 da agência, os beneficiários se distribuem da seguinte forma por modalidade (leva em consideração apenas operadoras médico-hospitalares<sup>1</sup>, foco principal desta pesquisa) (ANS, 2017)

Tabela 1 – Distribuição de Beneficiários por Modalidades de OPS – Mar/2017

<b>Modalidade</b>	<b>Quantidade de Beneficiários</b>	<b>Representatividade no setor (%)</b>
<b>COOPERATIVA MÉDICA</b>	17.633.093	37,04%
<b>MEDICINA DE GRUPO</b>	17.631.173	37,04%
<b>SEGURADORA ESPECIALIZADA EM SAÚDE</b>	6.507.298	13,67%
<b>AUTOGESTÃO</b>	4.837.229	10,16%
<b>FILANTROPIA</b>	997.548	2,10%

Fonte: Elaborada pelo autor com base nos dados do ANS TABNET (ANS, 2017).

A partir dos dados acima, constata-se que mais de 85% dos beneficiários de assistência médica concentram-se nas modalidades medicina de grupo, cooperativa médica e seguradora especializada em saúde, pode-se afirmar que o fato ocorre devido a essas operadoras, por sua natureza jurídica, não terem nenhum fator de seleção para admissão de seus beneficiários, sendo livre a adesão de planos nestas instituições com um único requisito principal: a contribuição financeira fornecida pelo beneficiário.

### 2.3 Estudos de Previsão de Insolvência e Indicadores Contábeis

A palavra insolvência não tem uma definição absoluta na literatura, enquanto Altman (1968) classificou como insolventes as empresas que entraram na justiça com pedido de concordata, Chung et al (2008) citam dois tipos de explicação para o termo que, segundo os autores pode ser entendido como a superioridade do passivo em relação ao ativo, o que não necessariamente reflete uma situação de insolvência no curto prazo, ou como a incapacidade de uma companhia de pagar suas dívidas antes do vencimento, contrapondo a definição anterior.

Independente do conceito escolhido, estudos sobre o tema são frequentemente desenvolvidos buscando sempre o refinamento dos modelos de previsão. A pesquisa pioneira a respeito do assunto foi de Fitzpatrick (1932) que experimentou avaliar indicadores de empresas falidas com a de não falidas por setor, chegando à conclusão de que os números apresentados pela segunda eram superiores aos da primeira e que a falência das corporações acontecia gradativamente.

<sup>1</sup> Operadoras médico-hospitalares: tipo de atenção prestado pelas operadoras oferecido por entidades ou empresas que operam planos médico-hospitalares ou médico-hospitalares e odontológicos, podendo oferecer, adicionalmente, prestação de serviços médico-hospitalares ou odontológicos a terceiros não contratantes do plano (ANS,2000).

Posteriormente, Beaver (1966) em seu estudo intitulado *Financial Ratios as Predictors of Failure* tentou identificar com abordagens empíricas quais indicadores financeiros possuíam maior relação com a previsão de insolvência, o quanto estes indicadores eram eficientes e qual a probabilidade de uma determinada empresa falir, visto que possuía algum dos indicadores selecionados. Para o estudo foram consideradas empresas industriais que faliram nos Estados Unidos no intervalo compreendido entre os anos de 1954 e 1964. O autor colheu informações de demonstrativos financeiros das empresas que negociavam suas ações na bolsa de valores norte americana e utilizou como base o período de 5 exercícios contábeis anteriores a falência.

Beaver (1966) reuniu 79 empresas falidas e 79 saudáveis, no entanto, devido à dificuldade de coletar os dados de empresas falidas durante o período estipulado, a amostra final possuía 117 empresas. Na montagem dos pares “falidas-não falidas” optou por utilizar empresas do mesmo setor acreditando que a montagem das duplas poderia influenciar os resultados da pesquisa. O autor definiu que seriam consideradas falidas empresas que não conseguiam honrar seus compromissos financeiros quando vencidos.

Para seleção dos indicadores, Beaver (1966) considerou a literatura de finanças calculando 30 índices para cada período sobretudo índices de fluxo de caixa, endividamento, rentabilidade entre outros. Para identificar quais indicadores previam com melhor assertividade a insolvência das empresas, foi calculada a média dos indicadores das empresas solventes e das falidas, no qual pode se observar que os indicadores da primeira têm pouca oscilação, enquanto os da segunda variam mais agressivamente. Nesta análise os índices de fluxo de caixa sobre dívida, retorno sobre ativos e endividamento sobre ativos foram o que demonstraram melhor performance na previsão de insolvência.

Na segunda parte do estudo, o autor montou modelos com cada um dos indicadores contanto quantas vezes um indicador acertava ou errava na classificação das empresas que passavam por dificuldade financeiras. O pesquisador concluiu que os indicadores financeiros podem ser utilizados na previsão de insolvência de empresas e recomendou que os próximos estudos sobre tema utilizassem indicadores financeiros de forma combinada.

Em 1968, Altman propôs um modelo chamado Escore Z, fruto da aplicação de análise discriminante múltipla no estudo *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Backruptcy*. O autor utilizou as demonstrações financeiras de uma amostra composta de 66 empresas, sendo 33 solventes e 33 insolventes, foram consideradas insolventes empresas que pediram falência entre os anos de 1946 e 1965, intervalo utilizado

no estudo. Altman (1968) utilizou 22 indicadores financeiros, dos quais cinco foram selecionados com base no nível de significância estatística formando o seguinte modelo:

**Quadro 1 – Fórmula discriminante de Altman (1968)**

$Z = 0,012X1 + 0,014X2 + 0,033X3 + 0,006X4 + 0,999X5$ <p>Onde:</p> <p>Z = Índice geral</p> <p>X1 = Capital de Giro / Ativo Total</p> <p>X2 = Lucros Retidos / Ativo Total</p> <p>X3 = Lucros Antes dos Juros e Imposto de Renda (LAJIR) / Ativo Total</p> <p>X4 = Valor de mercado do Patrimônio Líquido / Valor Contábil do Passivo Total</p> <p>X5 = Vendas / Ativo Total</p>
---

Fonte: Altman (1968), adaptado pelo autor.

O modelo proposto por Altman estabelecia que as empresas com escore Z inferiores a 1,81 tendem a falir e as empresas com o indicador acima de 2,99 tem baixa probabilidade de falência, as empresas que foram classificadas entre esses limites estão na chamada zona de penumbra, onde ocorreram erros de amostragem.

No cenário nacional, o primeiro destaque entre os estudos de insolvência foi o desenvolvido por Kanitz (1976), o autor se utilizou da análise de balanços para identificar que variáveis tinham um alto poder de explicação sobre a insolvência das empresas. O estudioso encontrou várias dificuldades na análise da amostra, uma vez que se utilizou de 516 indicadores extraídos das demonstrações financeiras, ainda que a literatura julgasse que alguns destes não tivessem relação direta com a saúde da empresa. A grande dificuldade em levantar estes dados se deu devido à falta de padronização entre os demonstrativos, entretanto, Kanitz concluiu que as informações contidas nestes relatórios foram úteis para prever a insolvência de empresas com um grau de assertividade elevado. O modelo proposto pelo autor, chamado de termômetro de Kanitz possuía, 5 variáveis, sendo 3 de liquidez em sua função discriminante, o modelo institui que empresas que possuíssem índices acima de 0 são solventes e, as abaixo de -3 são insolventes, enquanto as posicionadas entre 0 e -3 podem ser consideradas em estado de alerta.

O primeiro autor brasileiro a utilizar análise multivariada na previsão de insolvência de empresas foi Elizabetsky (1976) que utilizou análise discriminante para estudar a situação de solvência de bancos comerciais.

Castro (2003) aplicou a técnica de Redes Neurais Artificiais (RNA) para prever insolvência de empresas brasileiras, o autor selecionou uma amostra de empresas não financeiras de capital aberto, solventes e insolventes, com seus balanços publicados entre 1996 e 2002. A partir destes dados calculou indicadores contábeis sugeridos pela literatura, ainda que estes não tivessem relações diretas com o propósito da pesquisa. Posteriormente, aplicou a técnica para identificar quais destes tinham maior relação com a insolvência das

companhias. Os resultados do estudo evidenciam uma alta eficácia da utilização das RNA quando comparada a Regressão Logística e a Análise Discriminante também utilizadas no estudo. Todos os modelos construídos com a técnica apresentaram assertividade acima de 90% nas amostras de treinamento e validação.

Em busca de refinar os modelos de avaliação de risco de crédito no mercado de energia, Mori e Umezawa (2007) apresentaram um modelo de classificação utilizando *random forest*, o estudo se propunha a utilizar-se de uma nova abordagem na classificação do risco de crédito visto a baixa regulamentação e a competitividade do mercado de energia no Japão. O trabalho coletou e analisou dados de 19 companhias japonesas retornando um índice de classificação para cada uma delas representando o grau de risco em que se enquadrava. A técnica foi comparada com Árvores de Regressão e Classificação e Redes Neurais Artificiais sendo a melhor avaliada de acordo com os resultados, mostrando uma precisão na classificação maior em 10,53% do que a primeira e 5,26% do que a segunda.

Em 2009, Guimarães e Alves, se propuseram a calcular a probabilidade de insolvência de operadoras de saúde através de regressão logística, tomando como parâmetro que a situação de insolvência estava diretamente ligada ao patrimônio líquido das operadoras, assim foram consideradas insolventes todas as OPS com PL igual ou inferior a zero. O estudo utilizou uma amostra de demonstrações financeiras de 706 operadoras de saúde brasileiras no ano de 2004 e seu estado de solvência em 2015, foram calculados 17 índices sugeridos pela literatura como mais eficazes na previsão de insolvência, dentro os quais três foram escolhidos para compor o modelo final, conforme quadro a seguir:

Quadro 2 – Resumo do modelo proposto por Guimarães e Alves (2009)

$$\log\left(\frac{PI}{1-PI}\right) = -4,834 + 1,206 \times \log\left(\frac{PC}{PL}\right) - 5,716 \times \frac{LL}{AT} + 2,42 \times \frac{RT}{AT}$$

Onde:

AT = Ativo Total

LL = Lucro Líquido

PI = Probabilidade de Insolvência

PC = Passivo Circulante (excluindo-se as provisões)

PL = Patrimônio Líquido

RT = Receita Total (Líquida)

Valores sugeridos para corte:

Vermelho (alta probabilidade de insolvência): maior que 0,040

Amarelo (alguma probabilidade de insolvência): entre 0,025 e 0,040

Verde (insolvência improvável) menor que 0,025

Fonte: Guimarães e Alves (2009)

O teste para averiguar a eficácia do modelo foi feito utilizando os indicadores para o ano de 2005 e o estado de solvência das operadoras em 2006, revelando uma alta capacidade preditiva, enquanto o Escore Z de Altman (1968) tinha um indicador de precisão<sup>3</sup> de 36,22% para massa de treino e 68,77% para a massa de teste, Guimarães e Alves (2009) apresentaram 71,60% e 69,65% respectivamente.

### **3 METODOLOGIA**

Neste capítulo é exposto o tipo de pesquisa, todos os procedimentos de coleta e organização dos dados, os critérios de definição da amostragem, as técnicas estatísticas que serão utilizadas para a análise dos dados, e, por fim, a seleção dos indicadores contábeis a serem considerados para a pesquisa.

#### **3.1 Tipo de Pesquisa**

De acordo com Lakatos e Markoni (2003) esta pesquisa é documental, visto que se utiliza como base informações que constam em documentos. Tem por base dados secundários, que são de domínio público e fornecidos pela Agência Nacional de Saúde Suplementar. Estão disponíveis no endereço eletrônico da Agência e podem ser acessados de forma livre.

A pesquisa é ainda descritiva e quantitativa em sua natureza e abordagem do problema, uma vez que são empregados instrumentos estatísticos, tanto na coleta, como no tratamento e análise dos dados, com o objetivo de descrever a relação entre diferentes constructos (RICHARDSON, 2008)

#### **3.2 População e Amostra**

O foco deste estudo são Operadoras de Saúde Médico-Hospitalares. Considerou-se como população as Operadoras de Saúde que pertencem as modalidades Medicina de Grupo, Cooperativa Médica e Seguradoras Especializadas em Saúde que estavam ativas e tiveram suas demonstrações financeiras publicadas no site da ANS nos anos de 2014 a 2016. A modalidade de autogestão foi excluída da amostra por ser destinada ao uso específico de um determinado público e ter grande parte de suas despesas financiadas pelo ente patrocinador. As filantropias não foram estudadas, visto que não tem como fonte principal de receita a contribuição de seus beneficiários, tendo assim seus indicadores prejudicados, o que poderia potencialmente comprometer o resultado do estudo. Por fim, as Administradoras também foram excluídas da população pelo fato de não possuírem beneficiários e seu balanço patrimonial ter uma estrutura totalmente diferenciada dos demonstrativos das demais operadoras. Apesar das restrições, mais de 85% dos usuários de planos de saúde médico-hospitalares são contemplados com esta pesquisa, conforme exposto no capítulo 2 deste trabalho.

Foram então selecionadas quatro amostras, uma para a etapa de treino dos modelos, composta por 75 empresas, 25 insolventes e 50 solventes, e três para a etapa de validação, composta cada uma por 39 empresas, sendo 13 insolventes e 26 solventes. A



amostra de treino considerou o biênio 2014/2015 e as amostras de teste consideraram o biênio 2015/2016.

### 3.3 Coleta, Organização e Análise dos Dados

Os dados utilizados para o estudo foram coletados do site da Agência Nacional de Saúde Suplementar do Documento de Informações Periódicas das Operadoras de Planos de Assistência à Saúde – DIOPS. O demonstrativo é publicado trimestralmente, porém neste estudo foram utilizados exercícios contábeis completos, de janeiro a dezembro, dos anos de 2014 a 2016.

O conceito de insolvência adotado para esta pesquisa foi o utilizado por Guimarães e Alves (2009) que considera insolvente as Operadoras de Saúde que possuíam patrimônio líquido positivo em um determinado ano e este se tornou negativo no ano posterior.

Para a amostra de treino dos modelos identificaram-se as empresas que se tornaram insolventes em 2015 e o seu quadro de indicadores em 2014, destas foram retiradas 25 e para cada empresa insolvente selecionada, duas empresas solventes foram incluídas aleatoriamente na base. Para testar o modelo foram separadas 3 amostras, cada uma contendo 13 empresas insolventes no ano de 2016 e 26 insolventes juntamente com os indicadores de ambas em 2015. As empresas que apresentaram a maioria dos indicadores zeradas por falta de dados foram excluídas da população para não comprometer os resultados da pesquisa.

Os indicadores utilizados no estudo estão descritos no Quadro 3 e estão entre aqueles apontados por Soares (2006) como os melhores para avaliar o desempenho econômico-financeiro de Operadoras de Saúde, segundo revisão da literatura feito pela autora.

Quadro 3 – Indicadores Utilizados

Indicador	Descrição	Fórmula
CEnd	Composição do Endividamento	$PC / PC+ELP$
GAt	Giro do Ativo	$CE/AT$
ML	Margem Líquida	$RL/CE$
RAt	Rentabilidade do Ativo	$RL/AT$
LG	Liquidez Geral	$(AC + RLP)/(PC + ELP)$
IDA	Indicador de Despesa Assistencial	$EI/CE$
IComb	Indicador Combinado	$(EI+DA+DC) / CE$

Fonte: Elaborado pelo autor com base em Soares (2006)

Onde:

PC = Passivo circulante

ELP = Exigível a longo prazo

CE = Contraprestações efetivas

AT = Ativo total

RL = Resultado líquido  
 AC = Ativo circulante  
 RLP = Realizável a longo prazo  
 EI = Eventos indenizáveis  
 DA = Despesas administrativas  
 DC = Despesas de comercialização

Para analisar o grau de importância das variáveis em cada modelo foram utilizados métricas apropriadas de acordo com a técnica empregada. Na regressão logística observou-se o grau de significância. Para esta métrica utiliza-se o valor de 0,05 como referência, as variáveis que se posicionarem abaixo desta faixa podem ser consideradas importantes para o modelo visto que esta mede o quanto do resultado é explicado devido ao acaso e não pela variável em questão, ou seja, as variáveis que possuem significância abaixo de 0,05 tem probabilidade superior a 95% de estarem relacionadas com o resultado obtido. Para as Árvores de Regressão e Classificação, Random Forest, e Redes Neurais Artificiais os balizadores para o grau de importância foram a Impureza de Gini, o Erro quadrático Médio (MSE) e o Algoritmo de Garson, respectivamente. A importância da variável é diretamente proporcional ao valor de cada métrica, quanto maior, melhor.

### 3.4 Técnicas utilizadas

#### 3.4.1 Regressão Logística

A aplicação de Regressão Logística se mostra adere ao propósito deste estudo, uma vez que diferente da regressão linear, esta técnica tem uma variável binária como resposta representando sucesso ou fracasso que, para este caso específico, são a solvência ou insolvência de uma OPS. Hair *et al* (2005) afirmam que a técnica se mostra bastante robusta quando as hipóteses de normalidade e homocedasticidade não são atendidas sendo assim mais eficiente que a Análise Discriminante, na ausência dessas condições.

A Regressão Logística se utiliza de observações para estimar um modelo em forma de equação, os resultados do modelo para este estudo podem ser interpretados como a probabilidade das empresas se tornarem insolventes ou não. De acordo com Soares (2013), quando a variável resposta é binária a equação resultante da aplicação da técnica tem a seguinte estrutura:

$$E[Y] = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)} = [1 + \exp(-\beta_0 - \beta_1 X_1 - \dots - \beta_p X_p)]^{-1}$$

Em que  $Y$  é a variável resposta,  $X_1, \dots, X_p$  são as variáveis independentes e  $\beta_0, \dots, \beta_p$  são os coeficientes estimados a partir das observações por meio da máxima verossimilhança. A equação pode ser ainda linearizada e a função resposta *logit* expressa conforme abaixo:

$$\text{logit}[\pi(x)] = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

A técnica segue ainda alguns pressupostos, são eles: i) a contribuição de cada variável  $X_p$ ,  $p = 1, \dots, n$  é proporcional ao seu coeficiente  $\beta_p$ ; ii) a contribuição de uma variável independente é constante e não depende das outras variáveis independentes; iii) Os erros são independentes e apresentam distribuição binomial; iv) as variáveis independentes não são multicolineares; e v) a escala  $\text{logit}[\pi(x)]$  é aditiva e linear.

Guimarães e Alves (2009) comprovaram a eficácia da RL para previsão de insolvência de Operadoras de Saúde, o modelo final de seu estudo possuía três indicadores contábeis, sendo um de alavancagem, um de lucratividade e um operacional. A precisão na classificação de insolvência superou a de modelos como o proposto por Altman (1968).

A regressão logística foi empregada utilizando o software R, com os comandos *lrm()*, do pacote *rms*, *glm()*, do pacote *stats*, e *vif()*, do pacote *car*, que permitiram o ajuste do modelo, e a avaliação da qualidade do seu ajustamento e de possíveis problemas de multicolinearidade.

### 3.4.2 Árvores de Regressão e Classificação

O algoritmo CART (*Classification and Regression Trees*) foi proposto em Breiman et al. (1984) e trata-se de uma técnica não-paramétrica capaz de produzir tanto árvores de classificação, caso a variável dependente seja categórica, como de regressão, caso a variável dependente seja numérica (BASGALUPP, 2010).

Fonseca (1994) complementa afirmando que a grande vantagem deste algoritmo está na capacidade de pesquisar relações entre os dados ainda que elas não sejam evidentes e produzir resultados simples e legíveis expostos em forma de árvore de decisão.

As árvores geradas pelo algoritmo CART são sempre binárias, segundo Taconelli et al (2009) elas se baseiam em partições sucessivas de uma amostra, buscando constituir subamostras internamente homogêneas e a classificação dessas subamostras ocorre conforme alguma medida descritiva e predição de novos elementos, executada pela própria estrutura de classificação constituída. As árvores podem ser percorridas da sua raiz, primeiro nó de uma árvore que representa todo o conjunto de dados (SOARES; REBOUÇAS, 2014), até as folhas,

nós terminais, respondendo apenas a questões simples do tipo “sim” ou “não”. Adicionalmente, o CART dispõe de um tratamento especial para atributos ordenados e também permite a utilização de combinações lineares entre atributos (agrupamento de valores em vários conjuntos) (BASGALUPP, 2010). O CART expande a árvore exaustivamente, realizando pós-poda por meio da redução do fator custo-complexidade (BREIMAN et al., 1984). Segundo os autores, a técnica de poda utilizada é muito eficiente e produz árvores mais simples, precisas e com boa capacidade de generalização.

Visto que o problema deste trabalho é investigar quais variáveis tem maior relação com a insolvência de operadoras de saúde, o CART se apresenta como uma alternativa de alto valor. Segundo Timofeev (2004) quando o método é empregado para classificar, o algoritmo por si só indica quais variáveis são mais importantes, descartando as demais. Outra vantagem atribuída à CART é a capacidade de o resultado final não ser alterado ainda que sobre os dados sejam aplicados logaritmos ou radiciações. O método também consegue trabalhar de forma robusta com *outliers*, uma vez que isola essas observações em um nó a parte. (SOARES; REBOUÇAS 2014)

Foi utilizado o comando *rpart()* contido no pacote *rpart* para a aplicação da técnica. O próprio comando retorna a importância de cada variável através da importância de GINI. As árvores de classificação foram utilizadas na classificação de solvência das operadoras, visto que esta é uma etapa predecessora da construção do relacionamento entre o indicador contábil e a situação de solvência ou insolvência.

### **3.4.3 Random Forest**

Proposto por Breiman (2001), o *Random Forest* é um gerador de árvores de decisão de maneira aleatória, as árvores de decisão são geradas em diferentes subespaços amostrais obtidos por meio do método de amostragem bootstrap (NISBET; ELDER; MINER, 2009). Para Mori e Umezawa (2007) a técnica baseia-se no aprendizado conjunto da árvore de decisão e é capaz de lidar com muitos dados dimensionais. O RF mostra-se útil para medir a importância de uma variável a redução média de impureza (*average impurity reduction*) que mensura o efeito que determinada variável tem sobre a pureza dos nós gerados (Grömping, 2009). O método também apresenta uma forma alternativa de medir a importância das variáveis que se baseia no aumento do quadrado médio do erro (QME) quando os valores de uma variável são permutados (GENUER, POGGI; TULEAU-MALOT, 2010).

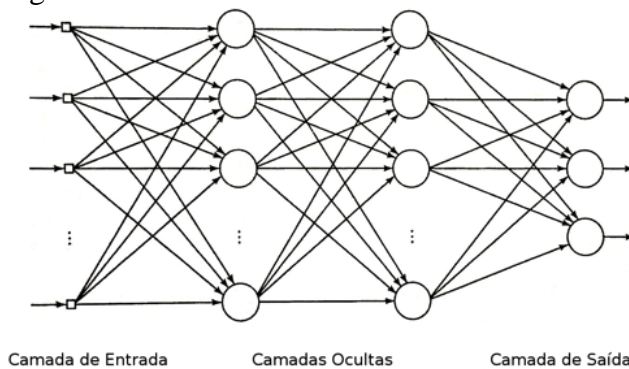
Neste estudo a técnica foi empregada através do comando `randomForest()`, que pode ser utilizado para classificação e regressão através da geração das árvores aleatórias, este

está contido no pacote *randomForest*. A importância de cada variável foi dada pelo comando *importance()* contido no mesmo pacote.

### 3.4.4 Redes Neurais Artificiais

O intuito das Redes Neurais Artificiais é identificar padrões e aprender comportamentos, ela possui uma estrutura de neurônios por onde a informação é processada, assimilada e reproduzida em forma de aprendizado. Para que o neurônio compreenda a função que deve exercer na classificação este passa por uma etapa de treinamento. O tipo de treinamento mais intuitivo utilizados nas ANN é o treinamento através de exemplos, por ele é apresentado à rede estímulos “x” que devem retornar saídas “y”. O processo de treinamento é repetido até que a rede consiga tornar a relação estímulo-saída, exposta acima, verdadeira. A camada mais externa da rede, primeiro neurônio, é chamada de camada de entrada, enquanto a última é nomeada camada de saída, as camadas intermediárias são chamadas de camadas ocultas (KOVACS, 2006). As redes com apenas uma camada, propostas por Rosenblatt (1957), são chamadas de Perceptron simples e estas só conseguem distinguir dados que sejam separáveis linearmente. Uma rede constituída por mais de uma camada é chamada de Perceptron de Multicamada (*multilayer perceptron* – MLP) que, por sua vez, consegue distinguir dados não-linearmente separáveis contanto que disponham de uma quantidade adequada de unidades em cada uma das camadas ocultas (SOARES, 2013).

Figura 1 - Estrutura de uma MLP



Fonte: Haykin (2001)

O Perceptron de Rosenblatt (1957) foi desenvolvido no intuito de resolver o problema de identificação de padrões. Esta tarefa é comum e trivial para os seres humanos que pode ser feita muitas das vezes de forma instantânea, porém, é um dos problemas mais complexos de serem resolvidos por uma máquina.

Neste estudo as Redes Neurais Artificiais propostas por Rosenblatt(1957) farão um papel fundamental ao observar o comportamento dos indicadores em determinado ano e verificar sua relação com a situação de solvência no ano posterior das Operadoras de Saúde.

A técnica foi utilizada através do comando *nnet()* disponível no pacote *nnet*. Para determinar a Importância relativa das variáveis de entrada na rede neural foi utilizado o comando *garson()* do pacote *NeuralNetTools*.

#### **3.4.4 Curvas ROC**

Para avaliar o desempenho dos modelos, foi utilizado o método das curvas ROC (*Receiver Operating Characteristic*). Esse tipo de análise baseia-se na relação entre a sensibilidade e a especificidade de classificadores binários. A sensibilidade é a taxa de verdadeiros positivos, que nesse estudo é a probabilidade de uma empresa solvente ser classificada como tal e a especificidade, é a taxa de verdadeiros negativos, a probabilidade de uma empresa insolvente ser classificada nesse grupo (SOARES, 2013). Tais probabilidades são computadas por meio da AUC (área abaixo da curva), principal estatística gerada por essa técnica e utilizada neste estudo para avaliar a capacidade preditiva dos modelos.

## 4 RESULTADOS

Neste capítulo serão apresentadas algumas estatísticas descritivas dos dados, as variáveis que mais se relacionam com a insolvência de operadoras de saúde e o desempenho das técnicas na previsão de insolvência das operadoras.

### 4.1 Estatísticas Descritivas

Antes de aplicar as técnicas de análise multivariada foram calculadas estatísticas descritivas em cima da amostra de treino, por ser a amostra com maior número de operadoras, exibidas na Tabela 2, no intuito de compreender sobre a dispersão dos dados. Utilizam-se os métodos de estatística descritiva para organizar, resumir e descrever os aspectos importantes de uma amostra observada (REIS; REIS, 2002).

Tabela 2 - Estatísticas descritivas dos dados

	Solventes			Insolventes		
	Mediana	Média	Desvio	Mediana	Média	Desvio
CEnd	0,7439	0,7124	0,1889	0,7606	0,7338	0,1722
GAt	1,8303	2,0189	0,6856	2,1800	2,5366	1,5037
ML	0,0160	0,0190	0,0559	0,0000	-0,0242	0,1198
RAt	0,0322	0,0333	0,0923	0,0000	-0,0757	0,2886
LC	1,4050	1,4808	0,7076	0,7129	0,6890	0,2845
LG	1,1090	1,2216	0,6875	0,6149	0,6234	0,2469
IDA	0,7796	0,7414	0,1522	0,8280	0,7965	0,1388
IComb	0,9488	1,0261	0,5989	1,0643	1,3472	1,0701

Fonte: Elaborada pelo autor

A princípio, nota-se uma dispersão semelhante nos indicadores das empresas solventes e insolventes, refletida pelo desvio padrão dos índices calculados. Pode ser observado Margem de Lucro (ML) e Retorno sobre o Ativo (RAt) negativos para a segunda massa. O Índice Combinado (IComb) revela uma situação delicada sobre as empresas insolventes, uma vez que a mediana deste indicador para o grupo é superior a 1. Ou seja, em pelo menos 50% das empresas insolventes, a cada 1 real recebido, 1,06 ou mais é pago.

### 4.2 Técnicas de Análise Multivariada

#### 4.2.1 Variáveis Predictoras de Insolvência

Aplicados os testes para verificação da importância das variáveis em cada modelo, obteve-se como resultado a Tabela 3.

Tabela 3 - Grau de importância das variáveis por técnica

RL		CART		RF		RNA	
Variável	Significância	Variável	GINI	Variável	MSE	Variável	Garson
LG	0,0002	LG	16,1093	LG	8,9596	LG	0,3592
IDA	0,1702	IComb	5,6032	RAt	5,7940	CEnd	0,1504
CEnd	0,1771	RAt	4,9028	ML	5,6091	IComb	0,1504
RAt	0,3193	ML	4,2024	IComb	5,5349	IDA	0,1329
IComb	0,3395	GAt	3,5020	GAt	3,4350	RAt	0,0929
GAt	0,4543	IDA	3,5020	IDA	2,0718	GAt	0,0835
ML	0,5748	CEnd	0,0000	CEnd	1,5112	ML	0,0361

Fonte: Elaborada pelo autor

Observou-se de forma unânime que a variável mais importante para a distinção entre OPS solventes e insolventes é a Liquidez Geral (LG). Na regressão logística, a significância desta se aproximou de zero, o que reflete um alto poder de explicação da variável no modelo. Já o índice de pureza de Gini para a LG foi quase o triplo quando comparada a variável posicionada como segunda colocada, enquanto o MSE também revela a maior contribuição desta variável para reduzir erros de classificação. Por fim, o algoritmo de Garson corrobora o que foi percebido pelos demais modelos. Outras variáveis podem ser destacadas como correlacionadas com a insolvência são o Índice Combinado (IComb) e o Retorno sobre o Ativo (RAt) que se posicionaram bem em mais de um dos modelos utilizados.

#### 4.2.2 Desempenho das Técnicas na Previsão de Insolvência

A aplicação das técnicas teve diferentes resultados nas amostras de teste e de treino dos modelos. Os resultados da amostra de treino estão dispostos na Tabela 4.

Tabela 4 - Classificação do Modelo - Amostra de Treino

Treino		Observações		Percentual de acerto geral	
		Insolventes	Solventes		
Classificação do método	RL	Insolventes	20 80,00%	4 8,00%	88,00%
		Solventes	5 20,00%	46 92,00%	
	CART	Insolventes	19 76,00%	4 8,00%	86,67%
		Solventes	6 24,00%	46 92,00%	
	RF	Insolventes	18 72,00%	7 14,00%	81,33%
		Solventes	7 28,00%	43 86,00%	
	RNA	Insolventes	24 96,00%	4 8,00%	93,33%
		Solventes	1 4,00%	46 92,00%	
Total		25	50		

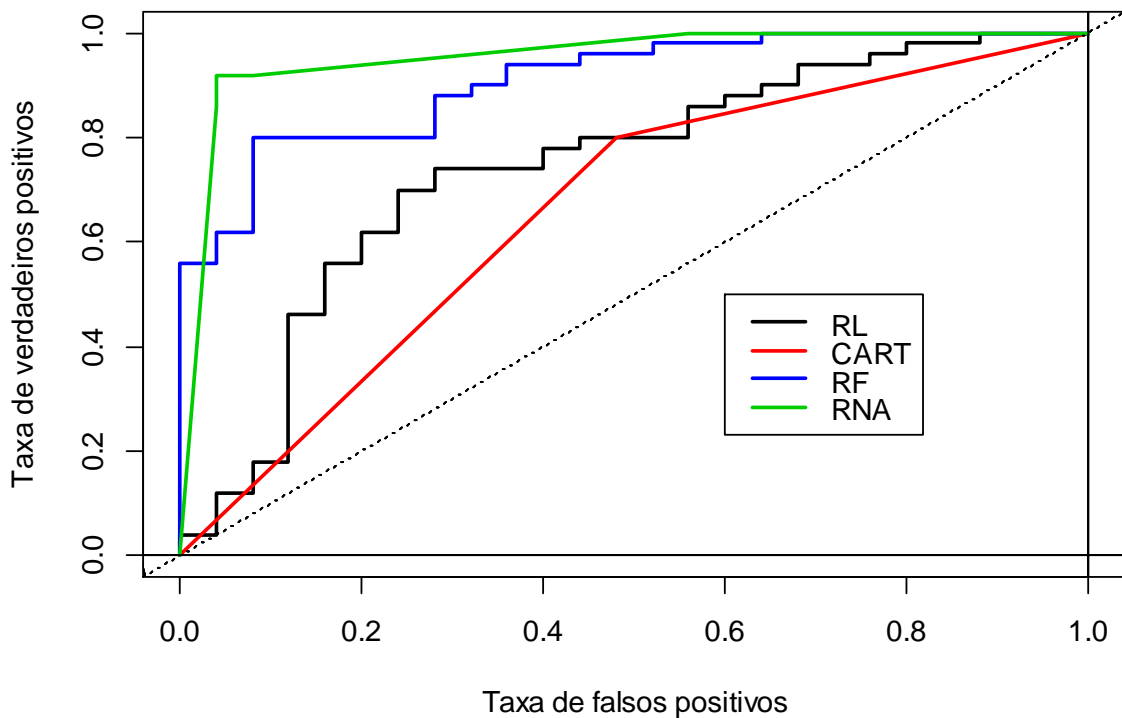
Fonte: Elaborada pelo autor



Para a amostra de treino a técnica que desempenhou melhor papel na previsão de insolvência foram as RNA com um percentual de assertividade de 93,33%, em contraposição, a técnica com menor eficiência preditiva foi a RF com 81,33% de precisão. Quando segregasse entre solventes e insolventes, há o destaque das RNAs na previsão de insolventes e empate entre RL, CART e RNA na previsão das empresas solventes. Apesar da diferença entre os modelos, pode se considerar que todos tiveram um desempenho satisfatório, haja vista que os percentuais de assertividade dos métodos ficaram acima de 80%.

As curvas ROC, construídas a partir da probabilidade de uma observação aleatória ser classificada corretamente, foram obtidas a partir da amostra de treino e estão disposta no Gráfico 2.

Gráfico 2 - Curvas ROC dos Modelos



Fonte: Elaborado pelo autor

Através do Gráfico 2 observa-se que as RNA obtiveram melhor desempenho na previsão de insolvência desta amostra. O que se comprova quando é observada a AUC de cada modelo. As RNAs obtiveram AUC de 0,9548, a segunda melhor posicionada foi a RF com 0,9080, logo abaixo ficou a RL cujo o indicador foi 0,7408 e o menor desempenho registrado foi o do CART atingindo o valor de 0,66.

Para fins de validação dos modelos gerados, todos foram testados com três amostras diferentes, montadas segundo a metodologia descrita anteriormente. Os resultados obtidos em cada um dos testes são apresentados nas Tabelas 5, 6 e 7.

Tabela 5 - Classificação do Modelo - Amostra de Teste 1

Teste 1		Observações		Percentual de acerto geral	
		Insolventes	Solventes		
Classificação do método	RL	Insolventes	10 76,92%	2 7,69%	87,18%
		Solventes	3 23,08%	24 92,31%	
	CART	Insolventes	7 53,85%	2 7,69%	79,49%
		Solventes	6 46,15%	24 92,31%	
	RF	Insolventes	12 92,31%	0 0,00%	97,44%
		Solventes	1 7,69%	26 100,00%	
	RNA	Insolventes	12 92,31%	2 7,69%	92,31%
		Solventes	1 7,69%	24 92,31%	
	Total		13	26	

Fonte: Elaborada pelo autor

O cenário do experimento para a primeira amostra de teste foi revertido em relação a de treino. A técnica que obteve maior destaque no segundo experimento foi a RF com 97,44% de acerto e a pior posicionada foi a CART com 79,49%. Novamente os modelos tiveram melhor performance na previsão de empresas solventes, com destaque para o RF que identificou empresas nesta situação sem erros.

Tabela 6 - Classificação do Modelo - Amostra de Teste 2

Teste 2		Observações		Percentual de acerto geral	
		Insolventes	Solventes		
Classificação do método	RL	Insolventes	10 76,92%	1 3,85%	89,74%
		Solventes	3 23,08%	25 96,15%	
	CART	Insolventes	7 53,85%	1 3,85%	82,05%
		Solventes	6 46,15%	25 96,15%	
	RF	Insolventes	12 92,31%	4 15,38%	87,18%
		Solventes	1 7,69%	22 84,62%	
	RNA	Insolventes	12 92,31%	4 15,38%	87,18%
		Solventes	1 7,69%	22 84,62%	
	Total		13	26	

Fonte: Elaborada pelo autor

Na segunda rodada de testes do modelo, obteve-se uma melhor performance da técnica de regressão logística com 89,74% de assertividade, a técnica com menor capacidade de predição para esta amostra foi o CART com 82,05%. Mais uma vez, os modelos mostraram melhor capacidade preditiva de empresa solventes e para esta amostra retomou-se o percentual de acerto maior que 80% para todos os métodos.

Tabela 7 - Classificação do Modelo - Amostra de Teste 3

Teste 3		Observações		Percentual de acerto geral	
		Insolventes	Solventes		
Classificação do método	RL	Insolventes	10 76,92%	1 3,85%	89,74%
		Solventes	3 23,08%	25 96,15%	
	CART	Insolventes	7 53,85%	1 3,85%	82,05%
		Solventes	6 46,15%	25 96,15%	
	RF	Insolventes	12 92,31%	2 7,69%	89,74%
		Solventes	1 7,69%	23 88,46%	
	RNA	Insolventes	12 92,31%	2 7,69%	92,31%
		Solventes	1 7,69%	24 92,31%	
Total		13	26		

Fonte: Elaborada pelo autor

Os modelos obtiveram suas melhores performances na última amostra de teste. As RNAs se destacaram com um percentual de acerto de 92,31% e novamente o CART foi o de menor performance com 82,05%. Na última amostra a predição de empresas solventes também foi a que obteve melhores resultados. Nesta rodada, a maioria das técnicas se aproximou do percentual de 90% de acerto.

De modo geral, observa-se na etapa de validação que as técnicas de *random forest* e de redes neurais artificiais obtiveram melhor desempenho de classificação, reforçando os resultados obtidos na etapa de treino. Além disso, com base nos resultados de todas as técnicas, foi possível elencar a Liquidez Geral como principal variável na segregação entre OPS solventes e insolventes, o que reforça a ideia de que as OPS precisam ter uma boa gestão de seus passivos frente aos seus ativos.

## 5. CONCLUSÃO

O objetivo geral deste estudo foi verificar quais indicadores estão relacionados com a situação de solvência ou insolvência de operadoras de saúde médico-hospitalares. Foram aplicadas quatro técnicas de análise multivariada, a partir disso o objetivo específico deste estudo é verificar qual das técnicas prevê com maior assertividade esse quadro, a fim de fornecer aos gestores destas companhias a oportunidade de traçar um plano de ação preventivo para que a insolvência não se concretize e aos beneficiários um acompanhamento mais próximo da situação financeira das empresas.

Foram utilizadas as técnicas de Regressão Logística, Árvores de Classificação e Regressão, *Random Forest* e Redes Neurais Artificiais aplicadas a 4 amostras de indicadores contábeis sugeridos pela literatura como os que melhor mediam o desempenho econômico-financeiro de Operadoras de Saúde. O estudo utilizou informações de demonstrações contábeis de Seguradoras Especializadas em Saúde, Medicinas de Grupo e Cooperativas Médicas dos anos de 2014 a 2016.

Por meio da aplicação das técnicas conclui-se que o indicador apresentado como o de maior relevância para segregar Operadoras de Saúde solventes e insolventes é a Liquidez Geral (LG) o que reforça a ideia de que as OPS precisam ter uma boa gestão de seus passivos frente aos seus ativos. Secundariamente, os indicadores Índice Combinado (IComb) e o Retorno sobre o Ativo (RA<sub>t</sub>) também tem uma relação com a insolvência destas instituições.

Quando são comparados os desempenhos das técnicas utilizadas na previsão da situação de insolvência das OPS, observa-se uma superioridade das Redes Neurais Artificiais frente as demais metodologias, conforme sugerido pela literatura. O método apresentou uma AUC de 0,9548 que representa a probabilidade de classificar corretamente uma observação aleatória, seguindo o critério da AUC pode-se considerar a performance do modelo como excelente.

Todas as técnicas apresentaram melhor desempenho na previsão de empresas solventes do que de insolventes, o que é comprovado pelo percentual de assertividade das primeiras que ficou muito próximo ou, na grande maioria das vezes, superior a 85% para todas as amostras de teste e a de treino.

Apesar do desempenho satisfatório dos modelos, eles não podem ser considerados o único balizador para determinar se uma Operadora de Saúde se tornará ou não insolvente, segundo a classificação sugerida neste estudo, no exercício posterior, visto os fatores externos

que influenciam na situação econômico-financeira destas instituições, tais como as exigências legais e a elevação dos custos assistenciais ano a ano.

Este estudo apresentou como limitação a ausência de informações de algumas operadoras, fato que determinou a exclusão destas instituições no trabalho. Outra limitação que pode ser pontuada é o fato da Agência Nacional de Saúde Suplementar (ANS) só divulgar as demonstrações contábeis das Operadoras em meados do exercício subsequente. Visto que neste estudo foram utilizados indicadores de um ano antes da entrada em insolvência, algumas empresas já tiveram sanções aplicadas pela agência que indicavam problemas no quadro econômico-financeiro antes mesmo das demonstrações contábeis terem sido publicadas.

Por fim, pode-se considerar que o estudo atendeu seus objetivos e para as futuras pesquisas, sugere-se a aplicação de novas técnicas estatísticas para previsão de insolvência de OPS. Sugere-se também estudar a relação entre as sanções aplicadas pela ANS às operadoras com sua situação econômico-financeiro. E, por último, propõe-se a aplicação destas técnicas em amostras separadas por porte, tipo de operadoras ou outro aspecto capaz de afetar a chance de insolvência.

## REFERÊNCIAS BIBLIOGRÁFICAS

ALTMAN, Edward I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Backruptcy. **Journal of Finance**, Boston, v.23, p. 586 - 609. 1968.

ALTMAN, E. I.; BAIDYA, T. K. N.; DIAS, L. M. R. Previsão de problemas financeiros em empresas. **Revista de Administração de Empresas**, Rio de Janeiro, v. 19, n. 1, p. 17-28, 1979.

BASGALUPP, Márcio Porto. **LEGAL-Tree: Um algoritmo genético multi-objetivo lexicográfico para indução de árvores de decisão**. Tese (Doutorado em Ciências da Computação e Matemática Computacional) – Universidade de São Paulo, São Paulo, 2010.

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**. Brasília, DF: Senado Federal: Centro Gráfico, 1988. 292 p.

BEAVER, William H. Financial Ratios as Predictors of Failure. **Journal of Accounting Research**, Chicago, Supplement of Selected Studies, p. 77-111. 1966.

BRASIL. Agência Nacional de Saúde Suplementar. **Demonstrações Contábeis**. Disponível em: <<http://www.ans.gov.br/perfil-do-setor/dados-e-indicadores-do-setor/demonstracoes-contabeis>>. Acesso em: 15/05/2017.

BRASIL. Agência Nacional de Saúde Suplementar. **Histórico**. Disponível em: <<http://www.ans.gov.br/aans/quem-somos/historico>>. Acesso em: 18/06/2017.

\_\_\_\_\_. Agência Nacional de Saúde Suplementar. **Caderno de Informação da Saúde Suplementar: beneficiários, operadoras e planos** – Rio de Janeiro, junho/2016. Disponível em: <[http://www.ans.gov.br/images/stories/Materiais\\_para\\_pesquisa/Perfil\\_setor/Caderno\\_informacao\\_saude\\_suplementar/caderno\\_JUNHO\\_2016\\_total.pdf](http://www.ans.gov.br/images/stories/Materiais_para_pesquisa/Perfil_setor/Caderno_informacao_saude_suplementar/caderno_JUNHO_2016_total.pdf)> Acesso em: 20/05/2017.

CAMPOS, Gastão Wagner de Sousa. **O público e o privado na saúde brasileira**. Le Monde Diplomatic, 2008.

CARVALHO, José Murilo de. **Cidadania no Brasil. O longo Caminho**. 3ª ed. Rio de Janeiro, 2002.

CASTRO, Francisco Henrique Figueiredo. **Previsão de insolvência de empresas brasileiras usando análise discriminante, regressão logística e redes neurais**. Dissertação (Mestrado) – Universidade de São Paulo, 2003.

\_\_\_\_\_. Congresso Nacional. **Lei n.º 9.656**, de 3 de junho de 1998. Publicada no Diário Oficial da União - D.O.U. em 04/06/98. Disponível em: <<http://www.ans.gov.br/portal/site/legislacao/legislacao.asp>> Acesso em: 18/01/2006.

\_\_\_\_\_. Congresso Nacional **Lei n.º 9.961**, de 28 de janeiro de 2000. Publicada no Diário Oficial da União - D.O.U. em 29/01/00. Disponível em: <<http://www.ans.gov.br/portal/site/legislacao/legislacao.asp>> Acesso: 18/06/2017.

DOMINGUEZ, Bruno. **Quando o público financia o privado - Quem ganha e quem paga a conta com as deduções de gastos com saúde?** Revista Radis, 2013.

ELISABETSKY, Roberto. **Um modelo matemático para decisões de crédito no Banco Comercial**. 1976. Dissertação (Mestrado) – Instituto Politécnico da Universidade de São Paulo, 1976.

FITZPATRICK, Paul J. **A Comparison of the Ratios of the Successful Industrial Enterprises with those of Failed Companies**. The Accountants Publishing Company, 1932.

GUIMARÃES, André L.S.; ALVES, Washington O. **Prevedo a insolvência de operadoras de planos de saúde**. São Paulo, 2009.

FONSECA, J. **Indução de árvores de decisão**. Dissertação (Mestrado) – Universidade de Lisboa, 1994.

KANITZ, Stephen Charles. **Como prever falências**. São Paulo, Mc Graw-Hill, 1978.

KOVACS, Zsolt L. **Redes Neurais Artificiais**. 4 ed. São Paulo: Livraria da Física, 2006.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Fundamentos de metodologia científica**. 5 ed. São Paulo: Atlas, 2003.

OHLSON, James A. Financial Ratios and the Probabilistic Prediction of Bankruptcy. **Journal of Accounting Research**, Chicago, v. 18, n. 1, 1980.

R Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Viena, Áustria, 2013.

RICHARDSON, Roberto Jarry. **Pesquisa social: métodos e técnicas**. 3 ed. São Paulo: Atlas, 2008.

SOARES, Maria Aparecida. THÓPHILO, Carlos Renato. CORRAR, Luiz João. **Avaliação de Indicadores Econômico-financeiro de Operadoras de Planos de Saúde Brasileiras: Uma aplicação da Análise Fatorial**. São Paulo, 2009.

SOARES, Rômulo Alves. **Modelos de classificação aplicados à previsão de insolvência de empresas brasileiras de capital aberto**. 2013. 87 f. Monografia - Curso de Ciências Atuariais, Universidade Federal do Ceará, Fortaleza, 2013. APÊNDICES

### APÊNDICE A – Operadoras Utilizadas no Estudo

REG_ANS	Razão Social	Solvente
355879	CAIXA DE ASSISTÊNCIA DOS ADVOGADOS DO ESTA	Não
351644	SMH - SOCIEDADE MÉDICO HOSPITALAR LTDA	Não
306428	ASSOCIAÇÃO BENEFICENTE DE CAMPO GRANDE	Não
343315	UNIMED MONTE CARMELO COOPERATIVA DE TRABAL	Não
346292	ASSOCIAÇÃO PORTUGUESA BENEFICENTE VASCO DA	Não
361852	UNIAO DE CLINICAS RIO GRANDE LTDA	Não
363286	UNIMED DE TAUBATÉ COOPERATIVA DE TRABALHO	Não
384356	COOPUS - COOPERATIVA DE USUÁRIOS DO SISTEM	Não
311961	UNIMED DE MANAUS COOP. DO TRABALHO MÉDICO	Não
384585	SANAMED - SAÚDE SANTO ANTONIO LTDA	Não
340146	ASSOCIAÇÃO AUXILIADORA DAS CLASSES LABORIO	Não
345270	UNIMED DO ABC - COOPERATIVA DE TRABALHO MÉ	Não
344800	ALVORECER - ASSOCIAÇÃO DE SOCORROS MÚTUOS	Não
410047	SAUDE - SISTEMA ASSISTENCIAL UNIFICADO DE	Não
312509	UNIMED DE PAULO AFONSO COOPERATIVA DE TRAB	Não
320111	SAÚDE SIM LTDA	Não
306665	UNIMED DO GUARUJÁ COOPERATIVA DE TRABALHO	Não
311405	SMS - ASSISTÊNCIA MÉDICA LTDA	Não
402851	MULTI SAÚDE - ASSISTÊNCIA MÉDICA E HOSPITA	Não
400190	SANTO ANDRÉ PLANOS DE ASSISTENCIA MÉDICA L	Não
314609	UNIÃO SAÚDE LTDA.	Não
393321	UNIMED-RIO COOPERATIVA DE TRABALHO MEDICO	Não
354554	MULTICLINICA SERVICOS DE SAUDE LTDA	Não
414581	UNIÃO MÉDICA - COOPERATIVA DE TRABALHO MÉD	Não
392391	HOSPITAL MARECHAL CÂNDIDO RONDON LTDA	Não
347361	UNIMED FEDERAÇÃO INTERFEDERATIVA DAS COOPE	Não
408883	CENTRO MÉDICO FÁTIMA LTDA	Não
413402	UNILIFE SAÚDE LTDA.	Não
356107	UNIMED CRUZEIRO - COOPERATIVA DE TRABALHO	Não
304158	UNIMED BOA VISTA COOPERATIVA DE TRABALHO M	Não
344150	UNIMED DE SERTAOZINHO - COOPERATIVA DE TR	Não
312282	INTERMEDICI PIRACICABA ASSISTENCIA MEDICA	Não
323993	UNIMED PETROPOLIS COOPERATIVA DE TRABALHO	Não
363774	UNIMED TERESÓPOLIS COOPERATIVA DE TRABALHO	Não
416711	ITAGUAÍ SAÚDE LTDA.	Não
333221	AMHPLA COOPERATIVA DE ASSISTENCIA MEDICA	Não
352055	MEDPORTO ASSISTÊNCIA MÉDICA LTDA.	Não
418854	ASSOCIAÇÃO SANTA CASA SAÚDE DE SOROCABA	Não
402770	CLIMOL CLINICA MÉDICA E ODONTOLÓGICA SÃO F	Sim
323357	UNIMED CABO FRIO COOPERATIVA TRABALHO MÉDI	Sim
362620	UNIMED SANTOS DUMONT SOCIEDADE COOPERATIVA	Sim
318477	OPERADORA UNICENTRAL DE PLANOS DE SAÚDE LT	Sim
403911	GOLDEN CROSS ASSISTENCIA INTERNACIONAL DE	Sim
335541	UNIMED VALE DAS ANTAS, RS - COOPERATIVA DE	Sim



418170	QUALLITY SAÚDE ASSISTÊNCIA MÉDICA AMBULATO	Sim
401196	SAÚDE GOIÂNIA LTDA	Sim
331872	UNIMED SAO JOSE DOS CAMPOS - COOPERATIVA D	Sim
355593	UNIMED PATROCÍNIO COOPERATIVA DE TRABALHO	Sim
367397	UNIMED CAMPINA GRANDE - COOPERATIVA DE TRA	Sim
477	SOMPO SAÚDE SEGUROS S.A.	Sim
326861	PROMÉDICA - PROTEÇÃO MEDICA A EMPRESAS S.A	Sim
369411	UNIMED DE ARACATUBA - COOPERATIVA DE TRABA	Sim
348830	PLENA SAÚDE LTDA	Sim
305472	UNIMED REGIONAL SUL GOIAS COOP. DE TRABALH	Sim
347507	UNIMED JI PARANÁ COOPERATIVA DE TRABALHO M	Sim
372561	UNIMED ALTO VALE - COOPERATIVA DE TRABALHO	Sim
312851	UNIMED CAMPO GRANDE MS COOPERATIVA DE TRAB	Sim
342131	UNIMED VILHENA COOPERATIVA DE TRABALHO MED	Sim
582	PORTO SEGURO - SEGURO SAÚDE S/A	Sim
350648	UNIMED REGIÃO DA CAMPANHA/RS - COOPERATIVA	Sim
352055	MEDPORTO ASSISTÊNCIA MÉDICA LTDA.	Sim
355691	UNIMED DO ESTADO DE SANTA CATARINA FED. ES	Sim
361615	UNIMED DE PARANAGUÁ COOPERATIVA DE TRABALH	Sim
371777	UNIMED NORTE CAPIXABA- COOPERATIVA DE TRAB	Sim
352683	UNIMED DE CAMPOS COOPERATIVA DE TRABALHO M	Sim
340251	UNIMED EXTREMO OESTE CATARINENSE COOPERATI	Sim
360414	UNIMED NOROESTE FLUMINENSE - COOPERATIVA D	Sim
364860	UNIMED DE TUBARAO - COOPERATIVA DE TRABALH	Sim
306762	UNIMED REGIONAL JAU - COOPERATIVA DE TRABA	Sim
334847	UNIMED DE JATAÍ COOPERATIVA DE TRABALHO MÉ	Sim
360244	PLANO DE SAÚDE ANA COSTA LTDA.	Sim
304344	UNIMED PARÁ DE MINAS COOPERATIVA DE TRABAL	Sim
330108	UNIMED CURVELO COOPERATIVA DE TRABALHO MÉD	Sim
354678	UNIMED MACHADO COOPERATIVA DE TRABALHO MED	Sim
328308	UNIMED DE MOCOCA COOPERATIVA DE TRAB. MÉDI	Sim
412228	POLICON ASSISTENCIA MÉDICA LTDA - EPP	Sim
367087	UNIMED/RS FEDERAÇÃO DAS COOPERATIVAS MÉDIC	Sim
413194	SANTA RITA SISTEMA DE SAUDE LTDA	Sim
348295	UNIMED SOROCABA COOPERATIVA DE TRABALHO MÉ	Sim
411868	EXTREMAMEDIC PLANOS DE SAÚDE LTDA	Sim
346209	UNIMED EXTREMO SUL COOPERATIVA DE TRABALHO	Sim
405051	CIMMAL OPERADORA DE PLANOS DE SAÚDE LTDA.	Sim
349712	UNIMED PONTA GROSSA COOPERATIVA DE TRABALH	Sim
333328	ASSOCIAÇÃO PRÓ-SAÚDE	Sim
309401	HOSPITAL NOSSA SENHORA DAS GRAÇAS LTDA.	Sim
348406	UNIMED DO RIO GRANDE DO NORTE - FEDERAÇÃO	Sim
379280	SICARD E SICARD ASSISTÊNCIA MÉDICA LTDA	Sim
413267	PREVENT SENIOR CORPORATE OPERADORA DE SAÚD	Sim
343811	SOCIEDADE BENEFICENTE UNIÃO OPERÁRIA DE AR	Sim
415081	PORTO DIAS SAUDE LTDA.	Sim

371564	UNIMED FRUTAL COOPERATIVA DE TRABALHO MEDI	Sim
352314	UNIMED JOÃO MONLEVADE COOPERATIVA DE TRABA	Sim
418731	ASSOCIAÇÃO DE SAÚDE DOS FORNECEDORES DE CA	Sim
412759	TERRAMAR ADMINISTRADORA DE PLANO DE SAUDE	Sim
358088	UNIMED LESTE PAULISTA COOPERATIVA DE TRABA	Sim
366561	BENSAUDE PLANO DE ASSISTENCIA MEDICA HOSPI	Sim
325236	ASSISTÊNCIA MÉDICA SÃO MIGUEL LTDA	Sim
415898	SOCIEDADE BENECAP DE ASSISTÊNCIA À SAÚDE	Sim
414298	ECOLE SERVIÇOS MÉDICOS LTDA	Sim
333689	MEDISERVICE OPERADORA DE PLANOS DE SAÚDE S	Sim
371629	UNIMED NOROESTE CAPIXABA COOPERATIVA DE TR	Sim
364584	UNIMED DE VOLTA REDONDA COOPERATIVA DE TRA	Sim
418676	CB SAÚDE ADMINISTRAÇÃO EM SAUDE SUPLEMENTA	Sim
419141	ASSOCIAÇÃO SÃO FRANCISCO VIDA	Sim
306126	UNIMED DE RIO CLARO SP COOPERATIVA DE TRAB	Sim
412015	MH VIDA - OPERADORA DE PLANOS DE SAÚDE - E	Sim
343064	GARANTIA DE SAÚDE LTDA	Sim
410888	FILOSANITAS SAUDE LTDA	Sim
416738	OESTE SAÚDE ASSISTÊNCIA A SAÚDE SUPLEMENTA	Sim
311294	UNIMED DE OURINHOS - COOPERATIVA DE TRABAL	Sim
417530	ASSOCIAÇÃO DE SAÚDE PORTUGUESA DE BENEFICÊ	Sim
314170	COOPERATIVA DE USUÁRIOS ASSISTÊNCIA MÉDICO	Sim
382540	ASSOCIAÇÃO BENEFICENTE DOS PROFESSORES PÚB	Sim
360147	CAMBORIÚ SAÚDE LTDA.	Sim
324566	UNIMED DE CATALÃO COOPERATIVA DE TRABALHO	Sim
312720	UNIMED DO ESTADO DO PARANÁ FEDERAÇÃO ESTAD	Sim
314099	UNIMED VALE DO SEPOTUBA - COOPERATIVA DE T	Sim
348520	MEDISANITAS BRASIL ASSISTÊNCIA INTEGRAL À	Sim
337188	COOPERATIVA DE TRABALHO MÉDICO DE POUSO AL	Sim
306886	UNIMED JUIZ DE FORA COOPERATIVA DE TRABALH	Sim
384577	UNIMED UBERLÂNDIA COOPERATIVA REGIONAL TRA	Sim
407011	GAMA SAUDE LTDA.	Sim
315265	PARANA ASSISTENCIA MEDICA LTDA	Sim
330264	UNIMED DE BARRA MANSA SOC. COOP. SERV.MED.	Sim
402834	UNIMED STA RITA, STA ROSA E SÃO SIMÃO COOP	Sim
354511	OPERADORA UNIESTE DE PLANOS DE SAÚDE LTDA	Sim
409464	AMENO ASSISTÊNCIA MÉDICA S/S LTDA.	Sim
325074	GREEN LINE SISTEMA DE SAÚDE S.A	Sim
402478	ORALCLASS ASSISTENCIA MÉDICA E ODONTOLOGIC	Sim
349755	SMEDSJ - SERVIÇOS MÉDICOS SÃO JOSE LTDA	Sim
323811	ASSISTENCIA MEDICO HOSPITALAR SAO LUCAS S/	Sim
392804	CENTRO CLÍNICO GAÚCHO LTDA	Sim
303739	SERMED-SAÚDE LTDA.	Sim
325465	AMPARA ASSISTÊNCIA MÉDICA PARAÍSO LTDA	Sim
356581	UNIMED ITAÚNA COOPERATIVA DE TRABALHO MÉDI	Sim
368253	HAPVIDA ASSISTENCIA MEDICA LTDA	Sim

327328	CASA DE SAÚDE NOSSA SENHORA DE FÁTIMA LTD	Sim
401081	AMESC - ASSOCIAÇÃO MÉDICA ESPÍRITA CRISTÃ	Sim
306959	UNIMED ALTO URUGUAI/RS - COOPERATIVA DE AS	Sim
322831	UNIMED ITAJUBA COOPERATIVA DE TRABALHO MED	Sim
309338	HOSPITAL REGIONAL DE FRANCA S/A	Sim
369233	UNIMED PALMEIRA DOS ÍNDIOS COOPERATIVA DE	Sim
343153	UNIMED SOUSA - COOPERATIVA DE TRABALHO MÉD	Sim
343013	CLIMEPE TOTAL LTDA	Sim
343684	UNIMED ALTO DA SERRA - SOCIEDADE COOPERATI	Sim
360481	VERA CRUZ ASSOCIAÇÃO DE SAÚDE	Sim
419753	MEDGOLD ASSISTENCIA MEDICA LTDA - ME	Sim
346276	UNIMED SALTO/ITU - COOPERATIVA MÉDICA	Sim
362832	UNIMED PLANALTO CENTRAL/RS - COOPERATIVA D	Sim
323268	UNIMED VALE DO SÃO FRANCISCO COOPERATIVA D	Sim
413721	SOCIEDADE BENEFICENTE DEZOITO DE JULHO	Sim
385255	UNIHOSP SAÚDE S.A.	Sim
318566	UNIMED METROPOLITANA DO AGRESTE - COOPERAT	Sim
395480	ESMALE ASSISTENCIA INTERNACIONAL DE SAUDE	Sim
316148	UNIMED POÇOS DE CALDAS - SOC. COOP. DE TRA	Sim
406805	COOPERATIVA DOS USUÁRIOS DE SERVIÇOS DE SA	Sim
366340	UNIMED DE CAPIVARI -COOPERATIVA DE TRABALH	Sim
414492	LIFE EMPRESARIAL SAÚDE LTDA.	Sim
415847	ASSISTÊNCIA MÉDICO-HOSPITALAR LTDA.	Sim
341941	HOSPITAL DE PRONTOCLINICA LTDA.	Sim

## APÊNDICE B – Códigos Utilizados no R

```
install.packages("rms")
install.packages("rpart")
install.packages("rpart.plot")
install.packages("nnet")
install.packages("randomForest")
install.packages("NeuralNetTools")
install.packages("ROCR")
```

```
require(rms)
require(rpart)
require(rpart.plot)
require(nnet)
require(randomForest)
require(NeuralNetTools)
require(ROCR)
```

```
set.seed(1)
```

```
##### Amostra de Treino 1 #####
```

```

amostra1<-read.csv("Amostra 1 - Treino v2.csv",header=T,sep=";",dec=",")
colnames(amostra1)
attach(amostra1)
summary(amostra1)
table(amostra1$Solvente)
equacao<-Solvente ~ CEnd + GAt + ML + RAt + LG + IDA + IComb

reg.log<-lrm(equacao,data=amostra1)
reg.log
log1<-glm(equacao,data=amostra1,family="binomial")
vif(log1)
summary(log1)

preditos.log<-predict(reg.log,type="fitted")
table(round(preditos.log),amostra1$Solvente)
prop.table(table(round(preditos.log),amostra1$Solvente),2)

cart<-rpart(equacao,data=amostra1,cp=0.00001)
rpart.plot(cart)
summary(cart)$variable.importance
preditos.cart<-predict(cart)[,2]
table(round(preditos.cart),amostra1$Solvente)

rf<-randomForest(equacao,data=amostra1,importance=T)
varImpPlot(rf)
importance(rf)[,4]
predict(rf,type="prob")

preditos.rf<-predict(rf,type="prob")[,2]
table(round(preditos.rf),amostra1$Solvente)

mod <- nnet(equacao, data = amostra1, size = 2)
plotnet(mod)
garson(mod,bar_plot=T)
colnames(amostra1)
preditos.rna<-predict(mod)
table(round(preditos.rna),amostra1$Solvente)

##### Análise de curvas ROC

# RL
roc.logit<-prediction(preditos.log,amostra1$Solvente)
performance.logit<-performance(roc.logit,measure="tpr",x.measure="fpr")
plot(performance.logit,lwd=2,xlab="Taxa de falsos positivos",ylab="Taxa de verdadeiros
positivos")
abline(h=0,v=1)
abline(a=0,b=1,lty=3)
auc.logit<-performance(roc.logit,measure="auc")

```

```
auc.logit
```

```
# CART
```

```
roc.cart<-prediction(preditos.cart,amostra1$Solvente)
performance.cart<-performance(roc.cart,measure="tpr",x.measure="fpr")
plot(performance.cart,lwd=2,xlab="Taxa de falsos positivos",ylab="Taxa de verdadeiros
positivos",add=T,col="red")
abline(h=0,v=1)
abline(a=0,b=1,lty=3)
auc.cart<-performance(roc.cart,measure="auc")
auc.cart
```

```
# RF
```

```
roc.rf<-prediction(preditos.rf,amostra1$Solvente)
performance.rf<-performance(roc.rf,measure="tpr",x.measure="fpr")
plot(performance.rf,lwd=2,xlab="Taxa de falsos positivos",ylab="Taxa de verdadeiros
positivos" ,add=T,col="blue")
abline(h=0,v=1)
abline(a=0,b=1,lty=3)
auc.rf<-performance(roc.rf,measure="auc")
auc.rf
```

```
# RNA
```

```
roc.rna<-prediction(preditos.rna,amostra1$Solvente)
performance.rna<-performance(roc.rna,measure="tpr",x.measure="fpr")
plot(performance.rna,lwd=2,xlab="Taxa de falsos positivos",ylab="Taxa de verdadeiros
positivos" ,add=T,col="51")

abline(h=0,v=1)
abline(a=0,b=1,lty=3)
auc.rna<-performance(roc.rna,measure="auc")
auc.rna
```

```
legend(.6,.5,c("RL", "CART", "RF", "RNA"),col=c("black", "red", "blue", "51"),lwd=c(2,2,2,2))
```

```
##### teste 1
```

```
teste1<-read.csv("Amostra 1 - Teste.csv",header=T,sep=";",dec=",")
```

```
preditos.teste.log<-predict(reg.log,newdata=teste1,type="fitted")
table(round(preditos.teste.log),teste1$Solvente)
prop.table(table(round(preditos.teste.log),teste1$Solvente),2)
```

```
preditos.teste.cart<-predict(cart,newdata=teste1)[,2]
table(round(preditos.teste.cart),teste1$Solvente)
```

```
preditos.teste.rf<-predict(rf,type="prob",newdata=teste1)[,2]  
table(round(preditos.teste.rf),teste1$Solvente)
```

```
preditos.teste.rna<-predict(mod,newdata=teste2)  
table(round(preditos.teste.rna),teste2$Solvente)
```

```
##### teste 2
```

```
teste2<-read.csv("Amostra 2 - Teste.csv",header=T,sep=";",dec=",")
```

```
preditos.teste.log<-predict(reg.log,newdata=teste2,type="fitted")  
table(round(preditos.teste.log),teste2$Solvente)
```

```
preditos.teste.cart<-predict(cart,newdata=teste2)[,2]  
table(round(preditos.teste.cart),teste2$Solvente)
```

```
preditos.teste.rf<-predict(rf,type="prob",newdata=teste2)[,2]  
table(round(preditos.teste.rf),teste2$Solvente)
```

```
preditos.teste.rna<-predict(mod,newdata=teste2)  
table(round(preditos.teste.rna),teste2$Solvente)
```

```
##### teste 3
```

```
teste3<-read.csv("Amostra 3 - Teste.csv",header=T,sep=";",dec=",")
```

```
preditos.teste.log<-predict(reg.log,newdata=teste3,type="fitted")  
table(round(preditos.teste.log),teste3$Solvente)
```

```
preditos.teste.cart<-predict(cart,newdata=teste3)[,2]  
table(round(preditos.teste.cart),teste3$Solvente)
```

```
preditos.teste.rf<-predict(rf,type="prob",newdata=teste3)[,2]  
table(round(preditos.teste.rf),teste3$Solvente)
```

```
preditos.teste.rna<-predict(mod,newdata=teste3)  
table(round(preditos.teste.rna),teste3$Solvente)
```