

APLICAÇÃO DE MODELOS DE MACHINE LEARNING PARA PREVISÃO DE INSOLVÊNCIA NO CONTEXTO BRASILEIRO

BRUNO MARCELINO BORGES DOS SANTOS

Universidade Federal de Minas Gerais

DR. POUERI DO CARMO MÁRIO

Universidade Federal de Minas Gerais

Resumo

O estudo teve como objetivo geral avaliar o desempenho de diferentes modelos de aprendizado de máquina para previsão de insolvência, de 51 indicadores contábeis de uma amostra 72 empresas brasileiras, comparando-os ao modelo de regressão logística. Foram utilizados modelos do tipo *random forest* e redes neurais artificiais (ANN), estimados por meio de validação cruzada, em uma base de dados em painel coletados entre 1 a 5 anos pré-insolvência, tratados com a metodologia de Análise de Componentes Principais. Os resultados obtidos sugerem que não há superioridade significativa entre modelos estatísticos e computacionais, na amostra de teste. Além disso, a base de dados também sinaliza que a relação linear mais significativa com a falência se dá por meio dos grupos de indicadores de liquidez e de lucratividade.

Palavras chave: Insolvência, Previsão, Aprendizado de Máquina.

1. Introdução

Segundo Altman (2002) e Beaver (1966), verifica-se que a possibilidade de realização de um diagnóstico preventivo da situação de solvência das empresas é de suma importância para diversos setores da economia em geral, por isso está sendo amplamente pesquisada (PEREIRA; MARTINS, 2015) para desde o estabelecimento de marcos regulatórios até a análise de risco de crédito a ser realizada por instituições financeiras, assim reduzindo o impacto negativo causado no sistema financeiro internacional (JONES; HENSHER, 2008). Diversos *stakeholders* confiam no sucesso de uma firma, e o risco de falência incorre em reduções drásticas no valor de mercado das companhias, o que pode afetar sua alocação de recursos e resultar em custos sociais elevados (NEOPHYTOU; CHARITOU; CHARALAMBOUS, 2001).

Tais fatores se tornam ainda mais relevantes em períodos onde há iminência de crises em larga escala - como a crise financeira asiática de 1997, Bolha PontoCom de 2000 e a crise financeira global de 2008 - que ainda impactam os mercados financeiros globais (JONES; JOHNSTONE; WILSON, 2017) e causaram a queda de grandes instituições sólidas que atuavam no mercado (JONES; HENSHER, 2008). Isso ocorre sobretudo em casos onde uma firma possui uma grande rede de influência sobre outras empresas, e pode gerar um impacto negativo em cascata no ambiente econômico internacional, causando desemprego e queda da qualidade de vida nos países onde a crise ocorre (BALCAEN; OOGHE, 2004). Dado o advento da globalização, as empresas estão cada vez mais interligadas e competitivas e os marcos regulatórios acentuados, o que torna o ambiente econômico cada vez mais vulnerável a situações de crise (ALTMAN; SAUNDERS, 1998).

Em contraposição aos estudos iniciais de Modigliani e Miller (1958) que assumem a eficiência de mercado e a separação entre decisões de investimento e financiamento, é importante ressaltar que o risco de falência pode afetar a capacidade de captação de recursos de uma empresa, o que reduz seu conjunto orçamentário de projetos viáveis e restringe sua criação de valor, definida como o valor presente líquido de seus projetos de investimento (BALCAEN; OOGHE, 2004).

Portanto, faz-se necessária a automação da estimação e implementação de tais modelos, frutos do progresso da sociedade na área de matemática, estatística, informática e inteligência artificial, a fim de aplicar regras quantitativas de decisão a uma ampla quantidade de empresas (BALCAEN; OOGHE, 2004) utilizando de dados publicamente disponíveis e técnicas estatísticas avançadas (DIDIER; DIGHAYE, 2002).

1.1 Objetivos Gerais e Específicos

Este estudo tem por objetivo a aplicação de modelos de previsão de insolvência - não pagamento de dívidas com os credores - para gerenciamento de risco no contexto brasileiro, em uma amostra de empresas solventes e insolventes. O trabalho será realizado com o respaldo de mais de 50 anos de estudos anteriores que se baseiam na utilização de indicadores contábeis como fonte de informação da situação financeira (risco) da empresa (MARIO; AQUINO, 2004).

Estudos feitos por Pereira e Martins (2015) a partir de mais de 200 pesquisas apontam que são comumente utilizados modelos estatísticos lineares - como Análise Discriminante Linear Múltipla (ADLM) - que têm pouca flexibilidade no sentido de assumirem uma grande quantidade de pressupostos rígidos para o comportamento das variáveis, ainda que sejam mais facilmente interpretáveis.

Junto a isso, sabe-se que a realização de previsões a partir de variáveis financeiras é uma tarefa complexa, logo é pertinente a comparação entre o ganho de precisão obtido a partir da utilização de modelos de machine learning não-lineares mais complexos e robustos como as Redes Neurais Artificiais (ANNs) e algoritmos de *boosting* em detrimento de modelos lineares mais simples como a Regressão Logística (OBERMANN; WAACK, 2015), por exemplo. Diversos fatores serão levados em consideração para avaliar se a perda da capacidade de generalização (ajuste perfeito à amostra de treinamento, i.e., *overfitting*) e o gasto computacional elevado incumbido na estimação dos modelos mais recentes compensam o ganho de informação obtido. Além disso, será avaliada a capacidade de aplicação de tais modelos no contexto brasileiro, de forma a avaliar se os resultados são realmente contundentes e se assemelham aos encontrados em estudos internacionais.

Diante do que foi exposto, o objetivo geral da pesquisa é de avaliar se o conjunto de indicadores coletados é representativo da situação de solvência das empresas. Espera-se, portanto, alcançar alguns objetivos específicos neste estudo, sendo eles:

- A) comparar os resultados de diferentes modelos para previsão de insolvência, em cinco bases de dados de indicadores contábeis, como forma de estabelecer maneiras confiáveis de se mensurar o risco de crédito;
- B) identificar relações entre os resultados obtidos e outros estudos relacionados;

- C) avaliar se modelos computacionais superam os modelos estatísticos em robustez e acurácia.

Espera-se dar um passo para que a comunidade científica possa desenvolver novas técnicas de previsão de risco de crédito que forneçam diagnósticos cada vez mais precisos sobre a situação financeira das empresas. Dessa forma, serão colhidos resultados positivos para a sociedade em geral, seja evitando a incorrência de custos de falência desnecessários por parte de acionistas e credores, auxiliando os gestores de risco na tomada de decisão, auxiliando os órgãos governamentais supervisores na manutenção da credibilidade do Sistema Financeiro Nacional, e assim por diante.

Na seção 2, o presente estudo apresenta uma revisão de literatura a respeito dos principais avanços científicos sobre o tema que ocorreram nos últimos 50 anos. Já na seção 3, é estruturada a metodologia e os processos computacionais utilizados para a transformação das variáveis e estimação dos modelos. Na seção 4, são apresentados os resultados e discussões a respeito deles, e por fim serão expostas as conclusões do estudo e potenciais sugestões para a realização de futuros trabalhos.

2. Revisão de Literatura

A falência corporativa é definida como a execução dos procedimentos falimentares previstos em lei, a partir do momento em que a empresa não detém mais capacidades de honrar suas dívidas com terceiros (condição de insolvência) e não é apresentada outra solução independente do sistema judiciário (MARIO; AQUINO, 2004). Segundo Mário (2002), a falência tem "efeito pedagógico" no mercado, à medida em que é a apresenta à sociedade os erros cometidos por gestores oriundos de má gestão empresarial. Além disso, Altman (1983) reforça que empresas menores são mais suscetíveis às pressões econômicas, e por isso têm mais chance de falir.

Um ponto importante a se destacar é de que o *default*, i.e., o não cumprimento de uma obrigação, não necessariamente implica em falência. Em algumas situações ele pode ocorrer por má administração do capital de giro da empresa, onde não há a necessidade da liquidação da empresa pelo fato de que seus benefícios para os credores não compensam os custos legais da condução de um processo de falência.

Diante disso, pode-se definir a insolvência como sendo um estado, o *default* como sendo um ato, e a falência como um processo judicial (MARIO; AQUINO, 2004).

2.1. Risco de insolvência e previsão com uso de indicadores contábeis

Mário (2004) faz uma clara distinção entre as linhas de pesquisa na área de falência. A partir dos trabalhos pioneiros de Beaver (1966) e Altman (1968) na área de previsão de insolvência (de interesse do credor) e estimativa de risco (de interesse do gestor), que é a linha à qual este estudo está vinculado, surgem outras linhas que abrangem tanto a relevância da informação contábil neste processo (KAPLAN; URWITZ, 1979) quanto no delineamento dos diferentes procedimentos incorridos no processo de falência, como custos de agência (HART, 2000).

Com relação à previsão de insolvência, Mário (2004) também ilustra que há diversas sublinhas de pesquisa. As mais importantes envolvem o uso de indicadores contábeis com técnicas de análise multivariada, e o uso de dados de mercado com a Teoria de Precificação Opções, que prevê que a empresa se torna inadimplente quando o valor de mercado de uma empresa (encontrado com base em suas opções emitidas) cai abaixo de certo nível. Uma

comparação entre as duas sub-linhas de pesquisa é feita por Hillegeist et al (2002), concluindo que os dados de mercado devem se aliar aos dados contábeis pois não refletem toda a informação disponível sobre a probabilidade de falência (MÁRIO, 2004).

Em primeiro lugar, a literatura relevante que antecede a linha de pesquisa de previsão de insolvência parte dos trabalhos desenvolvidos por Fitzpatrick (1932), que selecionou uma amostra de 19 empresas solventes e 19 empresas em processo falimentar para comparar quais indicadores (chamados “índices-padrões”) tinham maior potencial na previsão de falência, indicando que Dívida Líquida / Patrimônio Líquido era a melhor variável preditora. Entre os anos de 1930 e 1962, foram desenvolvidos diversos estudos pioneiros que avaliam a qualidade de certos indicadores financeiros na previsão de falência (PEREIRA; MARTINS, 2015).

A partir dos estudos realizados por Beaver (1966, 1968), com um modelo diferente para cada indicador e um grupo de 158 empresas consideradas metade insolventes e metade solventes, foi aplicada a Análise Discriminante Univariada. Beaver faz questão de ressaltar a diferença de importância entre erros tipo I (classificar erroneamente uma empresa insolvente) e tipo II (classificar erroneamente uma empresa solvente). Vale ressaltar que o autor deixa claro em suas pesquisas o fato de que se os indicadores fossem utilizados em conjunto, poderiam gerar melhores resultados.

A partir daí, Altman (1968) publica estudos de um conjunto de 5 indicadores preditores a partir de uma Análise Discriminante Linear Múltipla com o uso de softwares estatísticos, sendo 66 empresas de manufatura - metade solventes e metade falidas - com valores próximos de ativos totais. O modelo, conhecido como "z-Score de Altman", foi testado para cada ano antes da falência e apresentava superioridade na classificação de empresas para um ano antes do processo falimentar, mas deteriorava sua acurácia com o tempo. Posteriormente, Altman, Haldeman e Narayanan (1977) criaram o ZETA Score, alternativo ao modelo anterior de Altman (1968). O novo modelo tinha por objetivo realizar diversas alterações que tornam o uso da Análise Discriminante mais eficiente, aumentando a sua capacidade preditiva. Os ajustes envolviam a transformação da base de dados utilizada nas previsões, e a concentração em ramos específicos. Vale ressaltar que os trabalhos de Altman foram pioneiros e motivaram o surgimento de diversos novos estudos contemporâneos que testam modelos lineares como Logit e Probit, por exemplo, mas que não conduziram a resultados diferentes dos encontrados a partir de outras técnicas (PEREIRA; MARTINS, 2015).

A partir de 1990, com o desenvolvimento de softwares estatísticos que permitiam análises mais confiáveis, novos modelos foram desenvolvidos e aplicados na previsão de insolvência. Muitos deles exigem uma grande capacidade de processamento de dados, e dentre os mais populares, destaca-se o uso de Redes Neurais Artificiais (ALEXANDROPOULOS et al., 2019; AL-SHAYEA; EL-REFAE; ELL-ITTER, 2010) que variam conforme o método de otimização utilizado. Nota-se também a utilização de Máquinas de Vetores de Suporte (SVM) (SHIN TAIK SOO LEE, 2005), algoritmos de bagging (LIU et al., 2017) e algoritmos de *boosting* (CHEN; GUESTIN, 2016; NATEKIN; KNOLL, 2013) dentre os estudos mais populares. Dito isso, conclui-se que são várias as técnicas de modelagem matemática passíveis de serem utilizadas na previsão de insolvência. Podemos classificar os modelos possíveis em dois grandes grupos, que são compostos por técnicas de estatística multivariada (lineares) e algoritmos de aprendizado de máquina (não-lineares). O presente estudo visa explorar três tipos de modelos mais recorrentes com o objetivo de identificar a divergência na qualidade das previsões fornecidas em uma base de dados

brasileiros. São eles: Regressão Logística (linear), Random Forest (não-linear) e Redes Neurais Artificiais (não-linear).

2.2. Modelos preditivos em estudos anteriores

Segundo Karels (1987), modelos de previsão de insolvência são, por definição, equações matemáticas capazes de identificar empresas na iminência de serem liquidadas ou tornarem-se insolventes. Seu processo de estimação e análise se dá de forma semelhante aos demais modelos de *machine learning* utilizados para derivar hipóteses sobre o comportamento de uma variável (*target*) com base em resultados obtidos a partir de outras variáveis (*features*), sendo que os modelos mais complexos são estimados por meio de softwares computacionais (OBERMANN; WAACK, 2015).

A estimação de modelos preditivos se dá por meio de duas fases: estimação (treinamento) e teste. Na base de treinamento são ajustados os parâmetros ótimos - "padrões" na relação entre o comportamento do banco de dados e a variável dependente - que, aplicados às variáveis independentes com base no algoritmo do modelo escolhido, geram previsões para o comportamento da variável dependente. Já na base de teste, o algoritmo do modelo é aplicado em dados "desconhecidos" com o objetivo de avaliar a eficácia do modelo em prever dados do mundo real, ou seja, sua capacidade de generalização a outros cenários desconhecidos (MITCHELL, 1997).

2.2.1. Regressão Logística (Logit)

É um dos primeiros métodos estatísticos a serem utilizados em modelos de previsão de insolvência, após a Análise Discriminante (OHLSON, 1980).

Tem como objetivo construir a reta que melhor divide o conjunto de dados de maneira correta, transformando o resultado de uma regressão linear (a partir das variáveis explicativas) log-linear utilizando a função sigmóide (que enquadra todos os valores obtidos entre 0 e 1). Sendo assim, os coeficientes da regressão representam o aumento marginal na probabilidade de que $Y = 1$ (OBERMANN; WAACK, 2015).

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta X$$

Sendo que:

- $p/(1-p)$ = razão de probabilidade (*odd*) -> proporção entre a probabilidade de sucesso e a probabilidade de falha
- Cada coeficiente representa a variação percentual aproximada no valor do logaritmo da odd dada uma variação em seu elemento X associado em uma unidade.

Para a obtenção da probabilidade, faz-se a seguinte transformação:

$$p = \frac{e^{\beta_0 + \beta X}}{1 + e^{\beta_0 + \beta X}}$$

A forma mais comum de treinar uma regressão logística é selecionar o modelo que maximiza a função de verossimilhança. Ela estabelece uma recompensa com base em cada valor previsto por ele (valor da verossimilhança) e retorna um score final para cada tipo de

modelo estimado, que é obtido a partir do produto entre os valores das verossimilhanças por cada acerto/erro.

Vale ressaltar que esse score penaliza as previsões com base na distância entre o valor previsto e o real, como mostra a fórmula a seguir.

$$Verossimilhança = \begin{cases} p \\ 1 - p \end{cases}$$

Sendo p a probabilidade prevista de que $Y = 1$, a função retornará um valor p caso Y realmente seja 1 e 1-p caso Y seja 0, penalizando a distância entre o valor previsto e o valor real.

A regressão logística assume alguns pressupostos para o comportamento das variáveis, para realizar boas previsões. São eles:

- Existência de relação linear entre x e y (correlação positiva)
- Ausência de correlação entre as variáveis explicativas (multicolinearidade)
- Ausência de heterocedasticidade
- Resíduos apresentam média zero (não necessariamente seguem uma distribuição normal)

2.2.2. Redes Neurais Artificiais

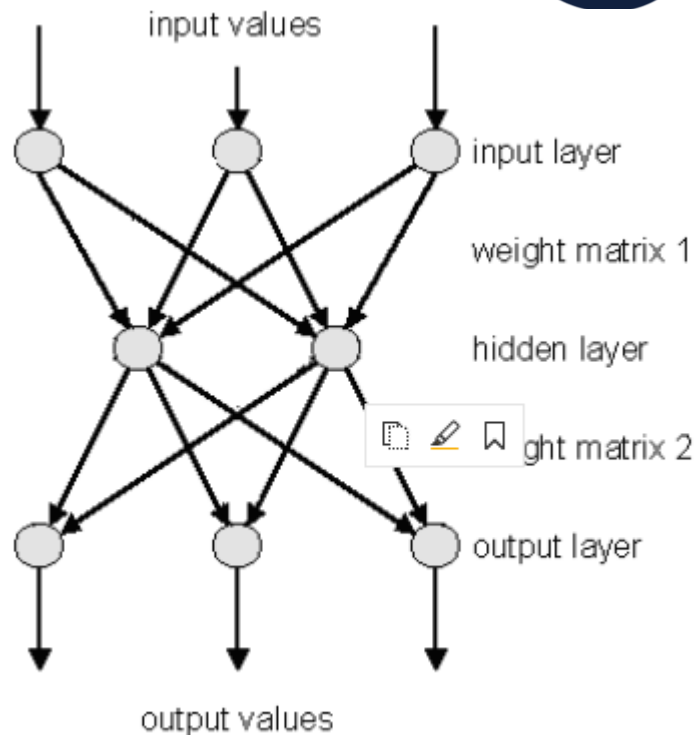
São modelos computacionais que funcionam de maneira semelhante ao cérebro humano, compostas por elementos altamente interconectados (neurônios) que respondem a estímulos (*features*), processam a informação por meio de funções (pesos) e repassam informações conjuntas a outros neurônios, até obter um resultado (AL-SHAYEA; ELREFAE; ELL-ITTER, 2010).

O modelo utilizado por este estudo em particular é do tipo mais simples de rede neural, conhecido como Perceptron Multicamadas (MLP). Nele, os neurônios funcionam como funções que recebem um vetor de inputs (x_1, x_2, \dots, x_n) e retornam um vetor de *outputs* (y_1, y_2, \dots, y_n), baseando-se em uma regra de cálculo específica estimada por meio de retropropagação.

No modelo MLP, cada neurônio realiza uma simples operação, que é definida como calcular uma função linear das variáveis (combinador linear) dado um viés (b) e pesos (w) para cada uma. Logo depois, aplicamos uma função de ativação em cima do resultado obtido de forma que ele seja um número entre 0 e 1. Também são funções de ativação conhecidas: ReLU, tangente hiperbólica (tanh), e a comumente utilizada função sigmóide. MLP é uma rede feed-forward, i.e, todos os neurônios enviam sinais para frente, em um único sentido.

$$y = f(w_1x_1 + w_2x_2 + \dots + w_nx_n + b) = \frac{1}{1 + e^{-(w_1x_1 + w_2x_2 + \dots + w_nx_n + b)}}$$

Esquema 1 - Funcionamento de uma rede neural



Fonte: Al-Shayea; Elrefae; Ell-Itter, 2010.

Para obter o melhor modelo possível, devemos alterar os valores dos pesos (w) e vieses (b) de cada neurônio de forma a obter o output desejado na maioria dos casos. Para isso, a rede neural MLP inicia esses coeficientes a partir de valores randômicos e tenta minimizar a função de custo (distância entre valor previsto e real) de forma iterativa, até chegar em um ponto ótimo. O número de neurônios e camadas escondidas deve ser fornecido previamente.

Um método comum utilizado para a realização desta tarefa é o Stochastic Gradient Descent, que se resume ao cálculo do gradiente da função de custo (vetor com as derivadas parciais da função de custo em relação a cada peso e cada viés). Sabemos que a direção apontada pelo gradiente de uma função indica a direção que maximiza seu valor naquele ponto, logo a partir do valor que está em direção oposta a ela podemos obter o valor que minimiza seu valor naquele ponto. Partindo desse princípio, podemos encontrar iterativamente o ponto cujos valores para os coeficientes minimizariam a função de custo e otimizaríamos a rede neural (ALEXANDROPOULOS et al., 2019).

2.2.3. Random Forest

Uma árvore de decisão é um mapa dos possíveis resultados de uma série de escolhas relacionadas, que podem ser usadas tanto para conduzir diálogos informais quanto para mapear um algoritmo que prevê a melhor escolha, matematicamente. Uma árvore de decisão geralmente começa com um único nó, que se divide em possíveis resultados. Cada um desses resultados leva a nós adicionais, que se ramificam em outras possibilidades. Assim, cria-se uma forma de árvore que permite realizar uma previsão para um conjunto de dados (LIU et al., 2017).

O ajuste de uma árvore de decisão ocorre de forma a partir da variável independente que apresenta o maior poder preditivo, subdividindo-se dessa maneira até encontrar uma previsão. Nesse contexto, o poder preditivo significa dividir com base em uma variável

independente (x), de forma que a variável dependente (y) possa ser separada em dois subconjuntos homogêneos bem definidos, obtendo assim o maior ganho de informação possível.

Uma divisão é feita dividindo a base atual em dois subconjuntos, sendo um direito (condição falsa) e o outro esquerdo (condição verdadeira). A qualidade de uma divisão pode ser medida por critérios de avaliação como a impureza de Gini e a Entropia, e é frequentemente comparada às demais possíveis divisões de forma a encontrar a melhor (BREIMAN, 2001).

O principal problema das árvores de decisão é de que elas tendem ao overfitting, ou seja, se adequar quase que perfeitamente ao conjunto de dados analisado. Isso significa que elas possuem alta variância, i.e., podem ser criadas árvores totalmente diferentes em cima de um outro subconjunto dos dados de treinamento (ainda que possuam o mesmo Processo Gerador de Dados).

Uma forma de reduzir isso é criar diversas árvores de decisão diferentes em cima de diversos subconjuntos dos dados de forma aleatória e com repetição (*bootstrapping*), selecionando inclusive as variáveis independentes de forma aleatória a fim de "descorrelacionar" os modelos ao máximo. A partir daí podemos analisar a resposta conjunta de diversos modelos, sendo este o cerne da criação das Random Forests.

A Random Forest, portanto, pode ser resumida como um modelo *ensemble* de várias árvores de decisão. Modelos *ensemble* constroem diversos modelos e analisam a sua resposta agregada como resultado da análise preditiva, e cada um possui suas peculiaridades. Em casos onde são utilizados subconjuntos aleatórios da base para treinamento, chamamos isto de *bootstrapping aggregation* (*bagging*) (LIU et al., 2017).

2.3. Processos de transformação de variáveis independentes

2.3.1. Análise de Componentes Principais

A Análise de Componentes Principais (também conhecida como Transformada de Hotelling ou Transformada Discreta de Karhunen-Loève) é uma forma de reduzir a dimensão do banco de dados, i.e, reduzir o número de variáveis a serem utilizadas para ajustar o modelo.

Para isso é aplicada uma transformação linear nos dados, a partir da qual são criadas novas variáveis com base nas antigas, chamadas de componentes principais (PC), que são combinações lineares das variáveis originais ponderadas pela sua variância e que são completamente descorrelacionadas entre si por definição.

A PCA também separa os componentes mais importantes para que estejam nas primeiras dimensões, nos eixos principais, com base no percentual da variância do banco de dados contido em cada eixo. Isso permite selecionar as melhores variáveis, que apresentam a maior parte do poder explicativo do banco de dados, e assim realizar previsões mais assertivas (MACKIEWICS; RATAJCZAK, 1993).

Para aplicar o método de PCA, deve ser realizada uma transformação que estabilize a variância dos dados, de forma que variações muito grandes de um indicador em termos

nominais não sejam levadas em consideração de forma semelhante a outro que varia pouco. Por isso, será aplicada a normalização em toda a base de dados, de forma a escalar os seus valores conforme a distância em relação à média em desvios-padrão.

2.3.2. Validação Cruzada

Para realizar a validação cruzada, dividimos a base de dados em “K” subconjuntos de treino. A ideia é de que sejam estimados “K” modelos diferentes e as métricas de avaliação do modelo final sejam calculadas como uma média simples das métricas obtidas para cada modelo de forma iterativa. O modelo final pode ser ajustado em cima de todo o conjunto de dados, de forma a obter resultados ainda mais precisos. Essa estratégia é cara computacionalmente, por isso é aplicada preferencialmente em conjuntos de dados pequenos. Este processo assume que as observações contidas no vetor de variáveis independentes sejam independentes e identicamente distribuídas. Além disso, a validação cruzada permite testar diversos conjuntos de parâmetros definidos *a priori* – também conhecidos como hiperparâmetros - ao estimar os modelos, permitindo encontrar o melhor modelo possível para ser aplicado em determinada base de dados.

2.4. Métricas de avaliação de resultados de modelos de classificação

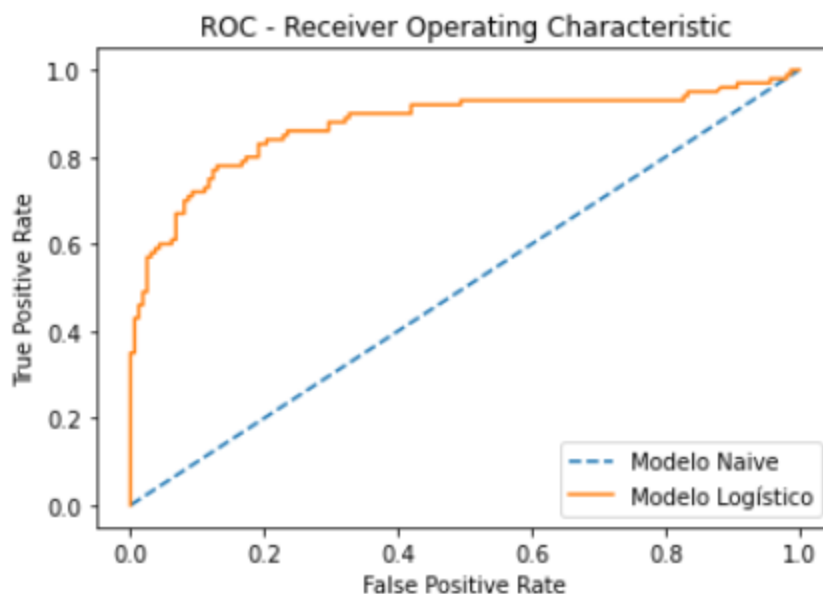
2.4.1. F1-Score

A estatística F1 calcula uma média harmônica entre duas estatísticas descritivas dos resultados de um modelo: a precisão e o *recall* (penalizando valores menores), a fim de maximizar ambos ao mesmo tempo. A precisão mede qual a taxa de acerto do modelo em suas previsões positivas, e o recall mede a taxa de acerto do modelo em suas previsões referentes aos valores que são positivos no conjunto de teste (FAWCETT, 2019).

$$F1 = 2 * \frac{(precision * recall)}{(precision + recall)}$$

2.4.2. Área Abaixo da Curva ROC (AUC)

A curva "*Receiver Operating Characteristic*" é uma técnica que permite a visualização e seleção de modelos classificadores que retornam probabilidades (entre 0 e 1) de um evento acontecer, com base em sua performance (FAWCETT, 2019). Ela pode ser calculada a partir de um mesmo modelo estimando-se a Taxa de Verdadeiros Positivos (sensibilidade) e a Taxa de Falsos Positivos (especificidade) para cada ponto de corte (probabilidade entre 0 e 1 a partir da qual o modelo passa a classificar os resultados como resultados positivos (1)), e pode servir para representar de outra forma esta relação de *trade-off*.

Gráfico 1 - Exemplo de curva ROC

Fonte: Elaborado pelo Autor.

A área abaixo da curva ROC (AUC), portanto, é uma estatística que descreve o comportamento da acurácia global do modelo no conjunto de teste (POLO; MIOT, 2020). Um valor maior para a AUC significa que, para cada ponto de corte possível, o modelo em geral apresenta maiores taxas de verdadeiros positivos e menor taxa de falsos positivos, o que indica a realização de previsões mais assertivas.

3. Metodologia

O presente estudo é caracterizado como pesquisa descritiva, no sentido em que permite analisar o fenômeno da previsão do risco de falência no contexto brasileiro a partir de procedimentos de caráter quantitativo - na medida em que se baseia em premissas estatísticas para a elaboração de modelos matemáticos e computacionais, com a interpretação de resultados - e levantamento bibliográfico a respeito do que já foi observado em estudos internacionais anteriores.

Os estudos de Pereira e Martins (2015) apontam que a estruturação de modelos de previsão de falência pode ser agrupada em dois aspectos: escolher quais indicadores financeiros utilizar nos testes empíricos, e o tipo de modelo a ser utilizado na previsão. O principal foco deste estudo reside na comparação entre técnicas estatísticas e computacionais de modelagem econométrica.

3.1. Coleta de dados

Serão coletados dados de demonstrações financeiras de 25 trios de empresas. Cada trio conterá empresas com duas características em comum: serem do mesmo setor e possuírem valores de ativos totais parecidos (JARDIN, 2015), sendo, portanto, uma empresa insolvente para cada duas empresas solventes. Somente serão coletados dados de empresas a partir de 2005, quando foi instituída a lei brasileira que regulamenta a recuperação judicial, a extrajudicial e a falência do empresário e da sociedade empresária, a fim de preservar a comparabilidade das demonstrações contábeis.

A análise utiliza de dados em painel, e para cada trio de empresas serão coletados os dados de 1, 2, 3, 4 e cinco anos de indicadores contábeis, antes da condição de insolvência da empresa insolvente que compõe o trio, por meio das plataformas Economatica e Refinitiv. Será realizada a estimação de modelos transversais, ou seja, para cada ano pré-falência serão filtrados os dados dos indicadores contábeis das empresas, e será estimado um modelo que objetiva prever uma variável dependente binária classificada como "0" caso a empresa seja solvente, e "1" caso a empresa tenha incorrido em processo de insolvência e possível liquidação dos seus ativos para o pagamento de credores.

Para cada ano pré-insolvência, haverá uma base de dados com 72 observações (empresas) e as variáveis coletadas para cada uma nesse corte temporal. Para estimar os modelos em suas respectivas bases, serão apagados os dados das empresas que contenham ao menos um dado faltante para qualquer indicador.

A escolha dos indicadores contábeis a serem utilizados como variáveis independentes no estudo será realizada aos moldes dos estudos de Jardim (2015), que se baseia nas dimensões jurídicas e conceituais a respeito do fenômeno de falência, observando liquidez (capacidade de saldar obrigações de curto prazo) e solvência (capacidade de pagar as obrigações já firmadas) a partir de grupos de índices de lucratividade, estrutura de capital, desempenho operacional e turnover.

É esperada a existência de correlação significativa entre os indicadores de liquidez e a variável *dummy* binária que representa a condição de insolvência das empresas, indicando que uma das principais causas do *default* é a má gestão de capital de giro das corporações. Indicadores de solvência e estrutura de capital também devem apresentar relação linear com o fenômeno de insolvência, dado que representam a influência da estrutura de financiamento da empresa em seu faturamento e sua proporção em relação à estrutura de capital. Os indicadores utilizados nos trabalhos de Jardim (2015) - e que foram utilizados neste estudo - estão elencados no Quadro 2.

Também é desejável que haja maior correlação entre os indicadores e a probabilidade de insolvência quanto mais próxima é a data na qual parte das empresas efetivamente se tornam insolventes, indicando que as variáveis apresentam maior poder preditivo em cenários mais críticos, onde é mais fácil distinguir uma empresa com dificuldades financeiras.

20^oCongresso USP de Iniciação
Científica em Contabilidade**Quadro 2** – Indicadores contábeis utilizados no presente estudo

Liquidez	Solvência	Lucratividade	Estrutura de Capital	Desempenho Operacional	Turnover
Disponível / Passivo Circulante (DN/PC)	Dívida financeira / Fluxo de Caixa Livre (DIV/FCL)	Fluxo de Caixa Livre / Patrimônio Líquido (FCL/PL)	Dívida LP / Patrimônio Líquido (DIVLP/PL)	Fluxo de Caixa Livre / Receita Líquida Operacional (FCL/RLO)	Contas a pagar / Receita Líquida Operacional (CAP/RLO)
Disponível / Receita Líquida Operacional (DN/RLO)	Despesas Financeiras / LAJIDA (DFIN/LAJIDA)	Fluxo de Caixa Livre / Ativo Total (FCL/AT)	Dívida LP/Ativo Total (DIVLP/AT)	Fluxo de Caixa Livre / Valor Adicionado (FCL/VA)	Ativo Circulante / Receita Líquida Operacional (AC/RLO)
Disponível / Ativo Circulante (DN/AC)	Despesas Financeiras / Lucro Líquido (DFIN/LL)	LAJIR / Patrimônio Líquido (LAJIR/PL)	Capital de Giro Operacional Líquido / Ativo Total (NOWC/AT)	LAJIR / Valor Adicionado (LAJIR/VA)	Estoque / Receita Líquida Operacional (EST/RLO)
Disponível / Ativo Total (DN/AT)	Despesas Financeiras / Ativo Total (DFIN/AT)	LAJIR / Ativo Total (LAJIR/AT)	Patrimônio Líquido / Passivo Exigível (PL/PE)	LAJIDA / Receita Líquida Operacional (LAJIDA/RLO)	Capital de Giro Operacional Líquido / Receita Líquida Operacional (NOWC/RLO)
Ativo Circulante / Passivo Circulante (AC/PC)	Despesas Financeiras / Valor Adicionado (DFIN/VA)	LAJIDA / Passivo Exigível (LAJIDA/PE)	Patrimônio Líquido / Ativo Total (PL/AT)	Receita Bruta/Receita Líquida Operacional (RB/RLO)	Contas a Receber / Receita Líquida Operacional (CAR/RLO)
Ativo Circulante / Ativo Total (AC/AT)		LAJIDA / Ativo Total (LAJIDA/AT)	Dívida Total / Patrimônio Líquido (DIV/PL)	Lucro Líquido / Receita Líquida Operacional (LL/RLO)	Receita Líquida Operacional / Ativo Total (RLO/AT)

20^oCongresso USP de Iniciação
Científica em Contabilidade**Quadro 2** – Indicadores contábeis utilizados no presente estudo (continuação)

Liquidez	Solvência	Lucratividade	Estrutura de Capital	Desempenho Operacional	Turnover
Passivo Circulante / Ativo Total (PC/AT)		Lucro Líquido / Patrimônio Líquido (LL/PL)	Dívida Total / Ativo Total (DIV/AT)	Lucro Líquido / Valor Adicionado (LL/VA)	
Passivo Circulante / Receita Líquida Operacional (PC/RLO)		Lucro Líquido / Ativo Total (LL/AT)		Valor Adicionado / Ativo Imobilizado (VA/IMOB)	
Estoque / Ativo Total (EST/AT)		LAIR / Patrimônio Líquido (LAIR/PL)		Valor Adicionado / Ativo Total (VA/AT)	
Ativo Circulante / Ativo Total (AC/AT)				Valor Adicionado / Receita Líquida Operacional (VA/RLO)	
Ativo Circulante / Passivo Circulante (AC/PC)					
Capital de Giro / Ativo Total (CDG/AT)					
Capital de Giro / Receita Líquida Operacional (CDG/RLO)					

Fonte: Elaborado pelo autor, 2022.

3.2. Modelagem

Os processos de manipulação das variáveis e geração dos modelos utilizados serão estimados por meio da linguagem de programação Python - conhecida no ramo da análise de dados - a partir das bibliotecas *sklearn* e *pandas*, já disponíveis e distribuídas com acesso gratuito para toda a comunidade internacional. Além disso, como representante dos modelos lineares clássicos convencionais será utilizada a regressão logística. Os resultados obtidos por meio de estatísticas descritivas dos resultados e erros do tipo I e tipo II serão comparados com dois exemplares de modelos não-lineares convencionais amplamente utilizados por cientistas de dados do mundo todo (ALEXANDROPOULOS et al., 2019; LIU et al., 2017), as *Random Forests* e as Redes Neurais Artificiais Multi-Camadas (MLP).

Tendo em vista que o banco de dados pode incorrer em fatores que podem comprometer a estimação dos modelos como a multicolinearidade, serão realizados procedimentos para sua transformação. Vale ressaltar que, ao contrário de outros modelos lineares tradicionais, a regressão logística não parte do pressuposto de normalidade das variáveis independentes para a sua estimação, sendo necessária somente a existência de relação linear entre cada variável independente e a variável dependente.

Para reduzir a multicolinearidade dos dados (descorrelacionando-os), será utilizada a Análise de Componentes Principais (PCA), que reduz a dimensão do conjunto de dados em pequenas componentes que prometem captar toda a sua variância. O método PCA permite gerar a mesma quantidade de variáveis contida na base original, porém em ordem crescente de poder explicativo. Por isso, foi selecionado somente o conjunto de variáveis que explica até 95% da variância dos dados, a fim de eliminar as componentes que possuem menor poder preditivo, e tornar o modelo mais robusto. Para realizar a PCA de forma correta, deve-se realizar uma transformação dos dados, de forma que as distribuições de todas as variáveis estejam em uma mesma escala. Por isso, foi realizada a normalização de todas as variáveis subtraindo todas as observações pela média de cada variável e dividindo-as pelo desvio-padrão, sendo que ambas as estatísticas encontradas são calculadas em cima das observações contidas na base de treinamento (estimação) do modelo.

Dada a quantidade reduzida de observações que serão coletadas, faz-se necessária a aplicação do método "K-Fold" de validação cruzada explicado por Cerqueira, Torgo e Mozetic (2019) para a divisão do conjunto entre bases randômicas de treino e teste, que servirão para estimar o modelo e avaliar os resultados respectivamente. A validação cruzada permite a redução do fenômeno do overfitting (ajuste completo à amostra de treinamento) em amostras pequenas. Será, portanto, aplicada a validação cruzada para três bases randômicas de treinamento e teste, e diversos modelos com conjuntos diferentes de hiperparâmetros definidos a priori. A validação cruzada será testada somente nos modelos de *Random Forest* e Redes Neurais, pois possuem diversos hiperparâmetros ajustáveis passíveis de otimização.

Quadro 3 – Hiperparâmetros Testados para a *Random Forest*

Hiperparâmetro	Valores Testados	Explicação
max_depth	0, 1 e 5	Profundidade máxima da árvore
min_samples_leaf	1, 3 e 5	Número mínimo de amostras presentes em cada folha (decisão)
min_samples_split	2 e 4	Número mínimo de amostras necessárias para fazer uma nova divisão
n_estimators	25, 50 e 100	Quantidade de árvores de decisão geradas
criterion	gini, entropy	Algoritmo que constrói as árvores de decisão

Fonte: Elaborado pelo autor, 2022.

Quadro 4 – Hiperparâmetros Testados para a Rede Neural

Hiperparâmetro	Valores Testados	Explicação
hidden_layer_sizes	(50,50), (50,100,50) e (100,0)	Quantidade de neurons por camada
activation	tanh e relu	Função de ativação de cada neurônio
solver	sgd, adam e lbfgs	Algoritmo que estima o modelo
alpha	0,0001 e 0,05	Peso para a regularização do tipo L2
learning_rate	constant e adaptive	Comportamento da taxa de otimização do modelo

Fonte: Elaborado pelo autor, 2022.

4. Resultados e Discussões

4.1 Análise dos Dados

Para realizar a estimação dos modelos, foram apagados os dados de empresas que possuem ao menos um dado faltante. A tabela 1 apresenta a quantidade de empresas solventes e insolventes contida em cada amostra coletada pré-insolvência. Vale ressaltar que a base de dados original continha 72 empresas, logo muitas delas tiveram que ser removidas para realizar a estimação dos modelos.

Tabela 1 – Quantidade de Observações por Ano Pré-Insolvência

Ano	Solventes	Insolventes	Total da Amostra
1	31	19	50
2	32	19	51
3	26	20	46
4	25	18	43
5	23	18	41

Fonte: Elaborado pelo autor, 2022.

O Apêndice A, por sua vez, contém estatísticas descritivas dos dados coletados para cada variável, dentro da amostra que foi efetivamente utilizada para a estimação e testagem dos modelos. Nota-se a existência de outliers em diversas observações, dada pela presença de valores extremos em determinadas contas que prejudicam a relação entre numerador e denominador à qual o indicador contábil representa. Outro fator relevante é a discrepância entre a magnitude da variância de determinados indicadores em relação aos demais. Isso pode vir a ter efeito prejudicial em modelos computacionais, que podem estabelecer peso muito maior para os indicadores que variam mais. Diante disso, torna-se relevante a realização da normalização de toda a base de dados, com o objetivo de estabilizar a variância de todos os indicadores, e analisá-los de forma proporcionalmente adequada.

Para demonstrar a relação entre cada grupo de indicadores contábeis e o fenômeno de insolvência, foi computada a correlação entre cada variável e a ocorrência ou não do fenômeno (variável binária). Os resultados obtidos sugerem que, em média, os indicadores de liquidez, lucratividade e estrutura de capital indicam maior correlação com o fenômeno de insolvência em relação aos demais. Isso indica que os indicadores que consideram a gestão do capital de giro e a margem de lucros da empresa são altamente relevantes no que diz respeito à previsão da saúde financeira da empresa.

Por outro lado, não é possível observar aumento na relação linear entre as variáveis e a condição de insolvência quanto mais próxima é a data onde ocorre o fenômeno. Isso se dá pela alta presença de ruído dada pela baixa quantidade de observações que foi possível coletar como amostra. Ainda assim, indicadores de liquidez e lucratividade se demonstram mais sólidos que os demais, em termos de poder explicativo. Os resultados foram apresentados na Tabela 2.

Tabela 2 – Correlação Média entre os Grupos de Indicadores e a Insolvência (em módulo)

Grupo	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Média
Liquidez	24,61%	19,84%	28,34%	20,92%	24,30%	23,60%
Solvência	4,81%	12,12%	9,43%	0,96%	20,90%	9,64%
Lucratividade	15,68%	26,68%	16,06%	32,27%	19,42%	22,02%
Estrutura de Capital	7,70%	21,67%	18,14%	19,22%	10,10%	15,37%
Desempenho Operacional	8,35%	9,60%	11,20%	6,93%	15,31%	6,44%
Turnover	1,65%	7,82%	15,46%	2,74%	2,90%	0,07%

Fonte: Elaborado pelo autor, 2022.

4.2. Apresentação dos Resultados Obtidos

As tabelas 3, 4 e 5 apresentam os resultados obtidos para as métricas de avaliação dos modelos de classificação que foram estimados. Destaca-se a superioridade dos modelos de Random Forest em relação aos demais, desde a avaliação do critério da área abaixo da curva ROC até a acurácia.

Outro fator relevante é que a Regressão Logística conseguiu resultados mais consistentes no tocante ao f1-score, ou seja, o balanceamento entre erros cometidos do tipo I e do tipo II. Isso significa que, ainda que seja um modelo estatístico, ele foi capaz de distinguir melhor as duas classes de empresas (solventes e insolventes), cometendo erros proporcionais para ambos os lados. O f1-score complementa a acurácia, no sentido de que, em amostras desbalanceadas, ele é capaz de avaliar a média ponderada de acertos para ambos os

lados, sendo uma métrica mais robusta que classifica melhor os modelos que não são enviesados a indicar somente um tipo de resultado.

Tabela 3 – Resultados da Regressão Logística

Ano	Acurácia	F1-Score	AUC
1	76,92%	66,67%	73,75%
2	61,54%	44,44%	59,52%
3	75,00%	57,14%	70,00%
4	90,91%	88,89%	92,86%
5	63,64%	60,00%	63,33%
Média	73,60%	63,43%	71,89%

Fonte: Elaborado pelo autor, 2022.

Tabela 4 – Resultados da *Random Forest*

Ano	Acurácia	F1-Score	AUC
1	84,62%	75,00%	80,00%
2	61,54%	28,57%	58,33%
3	75,00%	57,14%	70,00%
4	90,91%	88,89%	92,86%
5	63,64%	50,00%	61,67%
Média	75,14%	59,92%	72,57%

Fonte: Elaborado pelo autor, 2022.

Tabela 5 – Resultados da Rede Neural Artificial

Ano	Acurácia	F1-Score	AUC
1	76,92%	66,67%	73,75%
2	53,85%	40,00%	52,38%
3	75,00%	57,14%	70,00%
4	81,82%	80,00%	85,71%
5	54,55%	44,44%	53,33%
Média	68,43%	57,65%	67,03%

Fonte: Elaborado pelo autor, 2022.

4.3. Discussão dos Resultados

Com relação às premissas iniciais, verifica-se que o modelo de Redes Neurais é o menos adequado para ser utilizado nessa base de dados, pois além de representar um gasto computacional mais elevado que os demais, não apresentou melhores resultados que justifiquem sua aplicação, quando avaliado na base de teste. Isso pode ter sido causado por sua alta sensibilidade ao *overfitting* em bases de dados muito pequenas, ou aos pesos

aleatórios que são estabelecidos como premissas iniciais para se otimizar o modelo (NOVAK et al, 2018).

Verifica-se também que, em relação ao modelo Logístico, a Random Forest apresenta uma qualidade ligeiramente superior em suas previsões. Em contrapartida, é um modelo mais caro computacionalmente, sobretudo ao se realizar a otimização de seus hiperparâmetros. Dito isso, a escolha do melhor modelo depende da possibilidade de se renunciar a maior simplicidade em detrimento de resultados mais consistentes e robustos. Isso corrobora com os estudos de Obermann e Waack (2015), que concluem que não há diferença significativa entre modelos lineares e não-lineares.

Deve-se considerar também que a Regressão Logística exige a adoção de diversas transformações das variáveis utilizadas, para que elas atendam a seus pressupostos fundamentais. Em contrapartida, os modelos do tipo Random Forest e Redes Neurais requerem intervenção mínima na base de dados para sua estimação, além de serem mais robustos a outliers e dados faltantes, o que configura outra vantagem competitiva a ser considerada na tomada de decisão (JONES; HENSHER, 2008).

Ao analisar os resultados obtidos, o primeiro ponto a se destacar é o potencial efeito da baixa quantidade de dados para treinamento do modelo, na sua realização de previsões. Hair et al. (2009, p. 237) recomendam que, como regra prática, sejam utilizadas no mínimo 5 observações para cada variável inserida em um modelo logístico, como forma de estimá-lo de maneira robusta. No entanto, à medida em que não foi possível coletar a amostra de empresas necessária, são observadas diversas distorções como a ocorrência de previsões muito mais assertivas para os modelos Logit e Random Forest aos 4 anos antes da falência, por exemplo. Esse tipo de evento não era esperado pois a expectativa é de que os indicadores contábeis forneçam informações cada vez mais assertivas quanto mais próxima a empresa chega da condição de insolvência.

5. Conclusão

Foi realizado um estudo comparativo dos algoritmos mais utilizados para previsão de insolvência no contexto brasileiro, mediante pesquisa na literatura, com o modelo de Regressão Logística.

A fim de avaliar o momento em que os indicadores contábeis são mais eficazes em sinalizar potencial risco de insolvência, foram coletadas bases de dados de 1 a 5 anos que antecedem a condição de insolvência das empresas inadimplentes. Para balancear a amostra, foram coletados dados de duas empresas solventes para cada empresa insolvente, em mesmo período e com um tamanho do ativo similar.

Verifica-se que os indicadores de liquidez e lucratividade apresentam maior correlação com o fenômeno de insolvência, em relação aos demais. Por isso, uma limitação do estudo é de que análises mais robustas poderiam ter sido realizadas a fim de avaliar o seu poder preditivo em bases de dados brasileiras, em especial.

Além disso, o presente estudo também sugere que se pode obter previsões mais robustas com a utilização de bases de dados maiores para treinamento e teste dos modelos, porém não foi possível coletar dados suficientes devido às limitações inerentes ao contexto brasileiro. Ainda com relação à base de dados, poderia ter sido feita a suavização de outliers como forma de reduzir seus efeitos na qualidade das previsões fornecidas pelos modelos. No

Apêndice A, é possível verificar que há a presença de valores extremos na distribuição dos indicadores coletados, o que gera bastante ruído na amostra.

Também pode ser explorada uma maior variedade de algoritmos populares no mercado como a metodologia XGBoost (eXtensible Gradient Boost), outros tipos de redes neurais mais robustas como as Redes Neurais Convolucionais ou SOFM (self-organizing maps), máquinas de vetores de suporte, entre outros.

A principal contribuição deste estudo se dá por meio dos resultados obtidos, que sugerem que há superioridade do modelo de Random Forest em relação aos demais. As Redes Neurais não performaram tão bem nesta base de dados, porém a Regressão Logística obteve performance ligeiramente inferior à do melhor modelo. Por isso, a tomada de decisão sobre o modelo a ser escolhido na prática depende do trade-off entre flexibilidade e interpretabilidade. De um lado, a *Random Forest* é mais robusta a outliers e não necessita de transformações na base de dados para a sua implementação. Do outro, a Regressão Logística é mais facilmente interpretável e exige pouco gasto computacional para sua estimação, o que pode fazer com que os gestores sejam menos relutantes a utilizá-la na prática (OBERMANN; WAACK, 2015).

Este estudo conseguiu realizar uma análise coerente da linha de pesquisa sobre previsão de insolvência sob a ótica dos métodos computacionais mais avançados, que são amplamente utilizados pelo mercado, comparando-os com um modelo representativo das técnicas tradicionais.

6. Referências

- AL-SHAYEA, Q. K.; EL-REFAE, G. A.; ELL-ITTER, S. F. Neural networks in bank insolvency prediction. **Applied Science University of Jordan**, 2010.
- ALEXANDROPOULOS, S.-A. N. et al. A deep dense neural network for bankruptcy prediction. **ResearchGate**, 2019.
- ALTMAN, E. I. Financial ratios, discriminate analysis and the prediction of corporate bankruptcy. In: **The Journal of Finance**. 1968. p. 589–609.
- ALTMAN, E. I. Why business fail? **Journal of Business Strategy**, 1983. p.15–21.
- ALTMAN, E. I. Bankruptcy, credit risk, and high yield junk bonds. **Blackwell Publishers**, 2002.
- ALTMAN, E. I.; HALDEMAN, R.G.; NARAYANAN, P. ZETA analysis: a new model to identify bankruptcy risk of corporations. **Journal of Banking and Finance**, p. 29-54, 1977.
- ALTMAN, E. I.; SAUNDERS, A. Credit risk measurement: Developments over the last 20 years. **Journal of Banking and Finance**, p. 1721–1742, 1998.
- BALCAEN, S.; OOGHE, H. 35 years of studies on business failure: an overview of the classical statistical methodologies and their related problems. **Ghent University**, 2004.
- BEAVER, W. Financial ratios as predictors of failure, empirical research in accounting: Selected studies. In: RESEARCH, S. to Journal of A. (Ed.). **Journal of Business Finance Accounting**, p. 71–111, 1966.
- BEAVER, W. Alternative accounting measures as predictors of failure. In: REVIEW, T. A. (Ed.). **Journal of Business Finance Accounting**. p. 112–122. 1968.
- BREIMAN, L. Random Forests. **Machine Learning** 45, 5–32, 2001. Disponível em: <https://doi.org/10.1023/A:1010933404324>
- CERQUEIRA, V.; TORGO, L.; MOZETIC, I. Evaluating time series forecasting models - an empirical study on performance estimation methods. 2019. Disponível em: <https://doi.org/10.1007/s10994-020-05910-7>
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. 2016. Disponível em: <https://doi.org/10.1145/2939672.2939785>
- DIDIER, V. C.; DIGHAYE, A. The concept of “economic added result”, a new tool to prevent bankruptcy? **European Accounting Association Congress**, p. 1–30, 2002.
- FAWCETT, T. An introduction to roc analysis. **Elsevier**, 2019.
- FITZPATRICK, P. A comparison of tge ratios of the successful industrial enterprises with those of failed companies. **The Accountants Publishing Company**, 1932.
- HAIR, Joseph. et al. **Análise Multivariada de Dados**. 6th. ed. Porto Alegre: Bookma, 2009.
- HART, Oliver. Different Approaches to Bankruptcy. Discussion paper 1903, Harvard Institute of Economic Research. **Social Science Research Network**, 2000. Disponível em: http://papers.ssrn.com/paper.taf?abstract_id=241066
- HILLEGEIST, S. A. et al. Assessing the probability of bankruptcy. **Social Science Research Network**. 2002. Working papers. Disponível em: <www.ssrn.com>.
- JARDIN, P. du. Bankruptcy prediction using terminal failure processes. **Elsevier**, p. 286– 303, 2015.
- JONES, S.; HENSHER, D. A. Advances in credit risk modelling and corporate bankruptcy prediction. **Cambridge University Press**, 2008.

JONES, S.; JOHNSTONE, D.; WILSON, R. Predicting corporate bankruptcy: An evaluation of alternative statistical frameworks. **Journal of Business Finance Accounting**, 2017.

KAPLAN, R. S.; URWITZ, G. Statistical models of bond ratings: a methodological inquiry. **Journal of Business**, v. 52, n. 2, p. 231-261, 1979.

KARELS G., P. A. Multivariate normality and forecasting of business bankruptcy. In: SPRING (Ed.). **Journal of Business Finance Accounting**, p.573– 593, 1987.

LIU, C. et al. Generalising random forest parameter optimisation to include stability and cost, 2017. Disponível em: https://doi.org/10.1007/978-3-319-71273-4_9

MACKIEWICS, Andrzej; RATAJCZAK, Waldemar. Principal components analysis (PCA). **Computers & Geosciences**, v. 19, Issue 3, p. 303-342, 1993. Disponível em: [https://doi.org/10.1016/0098-3004\(93\)90090-R](https://doi.org/10.1016/0098-3004(93)90090-R)

MARIO, P. Contribuição ao estudo da solvência empresarial: uma análise de modelos de previsão – estudo exploratório aplicado em empresas mineiras. Tese (Doutorado) — **Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo**, 2002.

MARIO, P.; AQUINO, A. C. B. de. In: ATLAS (Ed.). **Teoria Avançada da Contabilidade**, p. 187–232, 2004.

MITCHELL, T. Machine learning. **McGraw Hill**, 1997.

NATEKIN, A.; KNOLL, A. Gradient boosting machines, a tutorial. **Frontiers in Neurorobotics**, 2013.

NEOPHYTOU, E.; CHARITOU, A.; CHARALAMBOUS, C. Predicting corporate failure: empirical evidence for the uk. working paper. **Department of Accounting and Management Science**, University of Southampton, p. 1–29, 2001.

OBERMANN, L.; WAACK, S. Demonstrating non-inferiority of easy interpretable methods for insolvency prediction. **Elsevier**, 2015.

OHLSON, J. A. Financial ratios and the probabilistic prediction of bankruptcy. **Journal of Accounting Research**, p. 109–131, 1980.

PEREIRA, V. S.; MARTINS, V. F. Estudos de previsão de falências - uma revisão das publicações internacionais e brasileiras de 1930 a 2015. **Revista Contemporânea de Contabilidade**, 2015.

POLO, T. C. F.; MIOT, H. A. Aplicações da curva roc em estudos clínicos e experimentais. **Scielo**, 2020.

SHIN TAIK SOO LEE, H.-j. K. K.-S. An application of support vector machines in bankruptcy prediction model. **College of Business Administration of Ewha Womans University**, 2005.

7. Apêndice A

Dados de 1 Ano Pré-Insolvência (50 amostras)

Indicadores	Média	Desvio-Padrão	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
DN/PC	0,49	0,74	0,00	0,03	0,18	0,70	4,42
DN/RLO	0,35	0,73	0,00	0,02	0,11	0,41	3,93
DN/AC	0,26	0,24	0,00	0,05	0,20	0,44	0,92
DN/AT	0,10	0,12	0,00	0,01	0,05	0,15	0,51
AC/PC	1,42	1,03	0,13	0,62	1,30	1,92	4,79
AC/AT	0,36	0,18	0,07	0,21	0,34	0,50	0,86
PC/AT	0,40	0,36	0,04	0,22	0,28	0,44	1,78
PC/RLO	1,58	4,09	0,03	0,36	0,59	0,93	28,31
EST/AT	0,09	0,11	0,00	0,02	0,07	0,14	0,65
AC/AT	0,27	0,14	0,03	0,16	0,26	0,33	0,70
AC/PC	1,11	0,87	0,02	0,45	0,97	1,60	4,76
CDG/AT	-0,04	0,41	-1,50	-0,12	0,08	0,24	0,42
CDG/RLO	-0,29	2,25	-14,55	-0,28	0,11	0,44	2,95
DIV/FCL	0,59	24,93	-100,36	-4,39	1,03	6,91	92,88
DFIN/LAJIDA	0,05	1,95	-9,20	-0,20	0,25	0,60	4,71
DFIN/LL	-3,62	31,74	-222,22	-0,73	0,13	0,98	16,73
DFIN/AT	0,09	0,14	0,00	0,03	0,05	0,09	0,90
DFIN/VA	-7,60	50,06	-353,79	-1,39	-0,15	0,37	6,59
FCL/PL	0,06	0,47	-1,46	-0,05	0,04	0,20	2,01
FCL/AT	-0,01	0,15	-0,38	-0,05	0,01	0,05	0,71
LAJIR/PL	0,14	0,53	-2,27	0,03	0,17	0,32	1,62
LAJIR/AT	0,01	0,23	-0,65	-0,03	0,04	0,08	1,16
LAJIDA/PE	0,47	1,13	-3,33	0,10	0,25	0,64	6,09
LAJIDA/AT	0,04	0,24	-0,63	-0,01	0,06	0,12	1,22
LL/PL	-0,15	1,32	-6,62	-0,01	0,07	0,19	2,08
LL/AT	-0,06	0,30	-1,13	-0,12	0,00	0,04	1,29
LAIR/PL	-0,14	1,32	-6,62	-0,05	0,10	0,23	2,10
DIVLP/PL	0,70	1,81	-1,19	0,04	0,44	1,00	12,30
DIVLP/AT	0,22	0,15	0,00	0,11	0,21	0,30	0,59
NOWC/AT	0,16	0,26	-0,47	0,05	0,17	0,34	0,70
PL/PE	1,92	3,14	-9,28	0,94	1,37	2,28	14,50
PL/AT	0,15	0,64	-2,04	0,14	0,32	0,44	1,83
DIV/PL	0,92	2,66	-3,16	0,11	0,72	1,29	17,11
DIV/AT	0,37	0,19	0,01	0,25	0,34	0,46	0,89
FCL/RLO	-0,17	1,10	-7,33	-0,11	0,01	0,10	1,46
FCL/VA	18,86	144,65	-113,46	-0,33	0,24	1,07	1014,70
LAJIR/VA	16,00	104,00	-12,75	-0,37	0,38	1,67	735,85
LAJIDA/RLO	-0,07	0,83	-4,74	-0,04	0,11	0,22	0,78
RB/RLO	0,25	0,22	-0,52	0,14	0,27	0,34	0,80
LL/RLO	-0,37	1,21	-6,63	-0,31	0,00	0,07	0,25
LL/VA	17,93	114,08	-8,42	-0,10	0,52	1,48	807,41
VA/IMOB	-2,63	13,90	-94,15	-0,34	-0,04	0,25	4,52
VA/AT	-0,09	0,45	-2,90	-0,09	-0,01	0,04	0,57



20º

Congresso USP de Iniciação
Científica em Contabilidade

VA/RLO	-0,03	0,28	-0,94	-0,19	-0,03	0,13	0,72
CAP/RLO	0,20	0,31	0,00	0,07	0,12	0,19	2,05
AC/RLO	1,29	2,17	0,01	0,36	0,64	1,21	13,76
EST/RLO	0,35	0,95	0,00	0,05	0,11	0,19	5,04
NOWC/RLO	0,77	1,80	-1,55	0,10	0,33	0,76	10,63
CAR/RLO	0,40	1,18	0,00	0,11	0,17	0,29	8,41
RLO/AT	0,73	0,89	0,03	0,31	0,51	0,85	5,11

Dados de 2 Anos Pré-Insolvência (51 amostras)

Indicadores	Média	Desvio-Padrão	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
DN/PC	0,55	0,97	0,00	0,03	0,32	0,57	6,16
DN/RLO	-0,30	5,66	-39,05	0,01	0,12	0,32	7,17
DN/AC	0,27	0,25	0,00	0,04	0,22	0,41	0,92
DN/AT	0,10	0,12	0,00	0,01	0,06	0,15	0,49
AC/PC	1,52	1,08	0,04	0,70	1,52	1,91	6,74
AC/AT	0,38	0,18	0,03	0,23	0,37	0,48	0,81
PC/AT	0,36	0,26	0,05	0,18	0,29	0,42	1,41
PC/RLO	-81,71	591,02	-4219,63	0,36	0,55	0,95	9,39
EST/AT	0,10	0,10	0,00	0,03	0,08	0,14	0,56
AC/AT	0,28	0,14	0,03	0,18	0,28	0,35	0,69
AC/PC	1,19	1,02	0,04	0,46	1,16	1,50	6,70
CDG/AT	0,02	0,32	-1,15	-0,06	0,10	0,24	0,34
CDG/RLO	79,60	566,44	-5,15	-0,10	0,21	0,61	4045,44
DIV/FCL	5,31	167,75	-949,38	-3,15	2,03	8,69	488,37
DFIN/LAJIDA	0,97	3,20	-4,93	0,08	0,27	0,96	17,95
DFIN/LL	0,70	5,99	-11,33	-0,79	0,08	1,01	34,71
DFIN/AT	0,07	0,08	0,00	0,02	0,04	0,09	0,33
DFIN/VA	2,74	18,74	-10,81	-0,53	0,14	0,74	129,56
FCL/PL	0,08	0,58	-1,24	-0,07	0,03	0,29	3,15
FCL/AT	-0,02	0,18	-0,72	-0,06	0,01	0,07	0,27
LAJIR/PL	0,13	0,47	-1,79	0,00	0,14	0,30	1,93
LAJIR/AT	0,00	0,15	-0,80	-0,01	0,04	0,07	0,16
LAJIDA/PE	0,32	1,18	-4,88	0,05	0,22	0,54	4,74
LAJIDA/AT	0,02	0,16	-0,77	0,02	0,05	0,10	0,22
LL/PL	0,02	0,65	-3,12	-0,07	0,05	0,14	2,55
LL/AT	-0,06	0,22	-1,40	-0,06	0,00	0,04	0,14
LAIR/PL	0,03	0,68	-3,13	-0,07	0,06	0,19	2,55
DIVLP/PL	0,57	0,97	-2,84	0,12	0,41	0,89	4,26
DIVLP/AT	0,21	0,15	0,00	0,09	0,18	0,34	0,60
NOWC/AT	0,20	0,20	-0,15	0,06	0,20	0,34	0,69
PL/PE	1,08	2,73	-8,46	0,73	1,23	2,00	7,83
PL/AT	0,17	0,55	-2,11	0,16	0,31	0,44	0,80
DIV/PL	1,05	1,95	-4,60	0,30	0,70	1,32	8,37
DIV/AT	0,36	0,18	0,02	0,24	0,35	0,46	0,91
FCL/RLO	9,31	69,96	-21,12	-0,09	0,02	0,13	498,65
FCL/VA	0,48	28,15	-150,23	-0,41	0,09	1,69	128,96



20º

Congresso USP de Iniciação
Científica em Contabilidade

LAJIR/VA	3,29	19,72	-6,46	-0,53	0,35	1,14	140,01
LAJIDA/RLO	13,65	97,80	-4,08	0,02	0,11	0,18	698,39
RB/RLO	0,91	4,62	-0,63	0,17	0,27	0,39	33,25
LL/RLO	24,37	175,71	-4,11	-0,08	0,01	0,07	1254,58
LL/VA	0,62	4,97	-17,42	-0,26	0,21	1,10	18,61
VA/IMOB	-0,52	4,14	-27,86	-0,40	0,01	0,30	4,71
VA/AT	-0,04	0,21	-1,17	-0,07	0,00	0,05	0,35
VA/RLO	1,32	9,44	-1,09	-0,14	0,02	0,14	67,41
CAP/RLO	-0,44	4,68	-33,10	0,06	0,10	0,21	2,36
AC/RLO	-2,11	24,69	-174,19	0,41	0,58	1,27	14,30
EST/RLO	0,34	1,39	-0,85	0,04	0,12	0,20	9,91
NOWC/RLO	0,20	5,17	-33,30	0,11	0,28	0,72	12,30
CAR/RLO	-0,86	8,38	-59,46	0,13	0,20	0,28	3,30
RLO/AT	0,65	0,58	0,00	0,28	0,57	0,83	2,88

Dados de 3 Anos Pré-Insolvência (46 amostras)

Indicadores	Média	Desvio-Padrão	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
DN/PC	0,37	0,42	0,00	0,03	0,23	0,50	1,42
DN/RLO	0,36	0,76	0,00	0,03	0,15	0,35	4,80
DN/AC	0,23	0,20	0,00	0,05	0,19	0,41	0,73
DN/AT	0,09	0,11	0,00	0,01	0,06	0,13	0,56
AC/PC	1,35	0,80	0,09	0,73	1,29	1,95	3,35
AC/AT	0,39	0,18	0,12	0,25	0,39	0,51	0,77
PC/AT	0,37	0,23	0,10	0,24	0,31	0,47	1,46
PC/RLO	5,13	21,29	0,09	0,40	0,50	1,03	133,78
EST/AT	0,11	0,09	0,00	0,04	0,09	0,15	0,48
AC/AT	0,29	0,15	0,06	0,17	0,26	0,37	0,69
AC/PC	1,00	0,60	0,04	0,50	0,96	1,35	2,53
CDG/AT	0,02	0,30	-1,33	-0,12	0,09	0,21	0,40
CDG/RLO	-2,10	14,86	-100,52	-0,14	0,14	0,41	2,48
DIV/FCL	65,57	327,04	-36,23	0,37	3,65	7,79	2194,60
DFIN/LAJIDA	0,45	1,02	-2,11	0,13	0,29	0,82	4,04
DFIN/LL	1,48	4,23	-5,79	-0,24	0,42	1,94	21,33
DFIN/AT	0,08	0,12	0,00	0,02	0,04	0,09	0,80
DFIN/VA	-0,17	5,66	-28,25	-0,54	-0,05	0,59	15,29
FCL/PL	0,00	1,25	-6,88	-0,09	0,07	0,29	3,92
FCL/AT	-0,02	0,28	-1,70	0,00	0,04	0,08	0,27
LAJIR/PL	0,04	2,14	-12,34	-0,01	0,17	0,33	6,54
LAJIR/AT	-0,02	0,44	-2,84	0,00	0,06	0,09	0,29
LAJIDA/PE	0,36	0,61	-2,00	0,08	0,39	0,57	1,73
LAJIDA/AT	0,00	0,44	-2,82	0,02	0,08	0,11	0,30
LL/PL	-0,04	2,57	-15,04	0,00	0,08	0,20	7,53
LL/AT	-0,08	0,50	-3,27	-0,04	0,02	0,05	0,32
LAIR/PL	-0,05	2,57	-15,08	-0,01	0,11	0,22	7,61
DIVLP/PL	0,51	1,01	-4,44	0,12	0,34	0,85	2,42
DIVLP/AT	0,20	0,16	0,00	0,06	0,14	0,35	0,60

**20º**Congresso USP de Iniciação
Científica em Contabilidade

NOWC/AT	0,22	0,20	-0,16	0,07	0,20	0,30	0,74
PL/PE	1,12	2,64	-9,01	0,79	1,16	1,93	11,94
PL/AT	0,22	0,44	-1,64	0,19	0,29	0,42	0,80
DIV/PL	1,03	2,28	-7,65	0,34	0,93	2,00	10,78
DIV/AT	0,36	0,19	0,02	0,22	0,36	0,45	1,01
FCL/RLO	-2,51	14,00	-94,40	0,00	0,07	0,12	0,54
FCL/VA	1,89	8,10	-17,61	-0,47	0,22	1,77	39,74
LAJIR/VA	1,13	7,79	-29,40	-0,71	0,47	2,12	39,05
LAJIDA/RLO	-0,44	2,97	-19,17	0,03	0,13	0,19	2,40
RB/RLO	0,24	0,19	-0,41	0,17	0,24	0,34	0,66
LL/RLO	-1,46	8,41	-56,88	-0,06	0,02	0,09	0,48
LL/VA	1,03	11,13	-33,86	-0,28	0,17	1,13	60,22
VA/IMOB	-0,56	4,46	-28,21	-0,43	-0,06	0,31	6,47
VA/AT	-0,04	0,17	-0,73	-0,08	-0,01	0,05	0,21
VA/RLO	-0,13	0,57	-2,67	-0,15	-0,04	0,12	0,49
CAP/RLO	0,81	4,41	0,02	0,06	0,11	0,23	30,06
AC/RLO	3,02	10,03	0,13	0,43	0,65	1,20	60,81
EST/RLO	0,44	1,15	0,00	0,08	0,14	0,24	6,28
NOWC/RLO	1,23	3,43	-0,32	0,12	0,30	0,85	21,89
CAR/RLO	1,39	7,00	0,03	0,12	0,21	0,32	47,72
RLO/AT	0,69	0,63	0,00	0,31	0,61	0,87	3,67

Dados de 4 Anos Pré-Insolvência (43 amostras)

Indicadores	Média	Desvio-Padrão	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
DN/PC	0,58	0,97	0,00	0,03	0,25	0,76	5,98
DN/RLO	17,80	100,23	0,00	0,03	0,19	0,37	653,30
DN/AC	0,26	0,23	0,00	0,06	0,20	0,44	0,84
DN/AT	0,11	0,12	0,00	0,02	0,08	0,15	0,57
AC/PC	1,74	1,77	0,10	0,80	1,29	2,03	9,08
AC/AT	0,40	0,21	0,05	0,21	0,37	0,56	0,77
PC/AT	0,33	0,24	0,06	0,18	0,27	0,41	1,35
PC/RLO	162,60	1033,29	0,03	0,38	0,48	0,85	6779,68
EST/AT	0,10	0,09	0,00	0,03	0,09	0,15	0,42
AC/AT	0,29	0,16	0,05	0,17	0,28	0,39	0,68
AC/PC	1,32	1,31	0,05	0,61	1,01	1,61	7,13
CDG/AT	0,07	0,31	-1,21	-0,04	0,09	0,27	0,67
CDG/RLO	-109,05	734,70	-4810,58	-0,12	0,12	0,43	187,51
DIV/FCL	-16,63	78,96	-498,92	-7,86	-0,97	4,53	28,23
DFIN/LAJIDA	-1,69	12,89	-80,35	0,07	0,27	0,68	13,31
DFIN/LL	-1,28	9,51	-45,04	-0,74	0,16	0,73	21,46
DFIN/AT	0,07	0,08	-0,04	0,02	0,04	0,07	0,38
DFIN/VA	-156,23	1115,01	-7281,26	-0,94	-0,01	0,47	454,35
FCL/PL	-0,09	0,46	-1,94	-0,23	0,01	0,10	1,06
FCL/AT	0,01	0,12	-0,35	-0,05	0,00	0,05	0,27
LAJIR/PL	0,08	0,37	-1,16	0,00	0,13	0,24	1,05
LAJIR/AT	0,04	0,08	-0,21	0,00	0,05	0,09	0,17

**20º**Congresso USP de Iniciação
Científica em Contabilidade

LAJIDA/PE	0,36	0,53	-0,66	0,05	0,34	0,49	2,07
LAJIDA/AT	0,06	0,09	-0,18	0,02	0,08	0,10	0,19
LL/PL	0,09	0,94	-2,28	-0,01	0,08	0,16	5,20
LL/AT	0,00	0,13	-0,41	-0,03	0,03	0,06	0,26
LAIR/PL	0,08	0,92	-2,09	-0,02	0,09	0,20	4,99
DIVLP/PL	0,53	1,02	-3,35	0,17	0,43	0,79	3,78
DIVLP/AT	0,21	0,17	0,00	0,05	0,17	0,30	0,66
NOWC/AT	0,24	0,24	-0,59	0,08	0,23	0,35	0,75
PL/PE	1,41	1,75	-5,50	0,83	1,33	2,01	7,17
PL/AT	0,28	0,46	-1,78	0,21	0,37	0,46	0,92
DIV/PL	0,70	2,59	-12,46	0,29	0,74	1,54	6,74
DIV/AT	0,35	0,20	0,02	0,23	0,34	0,46	1,11
FCL/RLO	93,90	606,96	-17,56	-0,10	0,00	0,09	3980,57
FCL/VA	-1376,86	9161,90	-60054,74	-0,70	-0,18	1,54	648,53
LAJIR/VA	-2442,77	16102,13	-105575,00	-0,36	0,41	1,73	443,69
LAJIDA/RLO	162,30	1067,39	-26,21	0,04	0,13	0,19	6998,83
RB/RLO	0,29	0,20	-0,35	0,18	0,27	0,37	1,00
LL/RLO	161,02	1061,68	-33,67	-0,03	0,04	0,11	6960,95
LL/VA	-2436,92	16016,23	-105019,57	-0,28	0,41	1,51	182,05
VA/IMOB	-0,68	4,66	-29,57	-0,26	-0,01	0,25	3,93
VA/AT	-0,02	0,16	-0,77	-0,05	0,00	0,03	0,21
VA/RLO	-0,10	0,71	-4,44	-0,15	-0,01	0,11	0,56
CAP/RLO	1,70	7,19	0,01	0,05	0,10	0,18	34,75
AC/RLO	53,55	301,95	0,13	0,41	0,60	1,17	1969,10
EST/RLO	2,33	13,47	0,00	0,06	0,13	0,27	88,47
NOWC/RLO	30,06	161,24	-0,84	0,17	0,33	0,98	1040,69
CAR/RLO	15,23	86,30	0,03	0,13	0,21	0,26	562,52
RLO/AT	0,67	0,67	0,00	0,27	0,50	0,89	3,86

Dados de 5 Anos Pré-Insolvência (41 amostras)

Indicadores	Média	Desvio-Padrão	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
DN/PC	0,44	0,52	0,00	0,07	0,20	0,65	2,00
DN/RLO	17,60	109,37	0,00	0,04	0,19	0,46	700,79
DN/AC	0,25	0,21	0,00	0,07	0,18	0,40	0,69
DN/AT	0,10	0,11	0,00	0,02	0,05	0,14	0,42
AC/PC	1,51	0,98	0,17	0,84	1,21	2,02	5,03
AC/AT	0,40	0,20	0,04	0,21	0,38	0,60	0,74
PC/AT	0,31	0,17	0,11	0,21	0,26	0,39	0,83
PC/RLO	95,73	579,59	0,05	0,37	0,55	0,91	3714,51
EST/AT	0,10	0,08	0,00	0,02	0,09	0,14	0,33
AC/AT	0,30	0,16	0,04	0,16	0,30	0,42	0,57
AC/PC	1,16	0,77	0,06	0,60	1,06	1,35	3,48
CDG/AT	0,09	0,22	-0,69	-0,06	0,06	0,26	0,45
CDG/RLO	-58,71	380,88	-2435,95	-0,09	0,15	0,49	80,32
DIV/FCL	4,53	16,64	-24,07	-0,84	2,28	6,33	95,43
DFIN/LAJIDA	1,28	3,71	-1,35	0,18	0,38	0,71	21,44



20º

Congresso USP de Iniciação
Científica em Contabilidade

DFIN/LL	8,03	31,87	-12,19	-0,47	0,36	1,04	157,72
DFIN/AT	0,06	0,06	0,00	0,02	0,04	0,07	0,29
DFIN/VA	8,83	49,79	-6,15	-0,89	0,00	0,45	315,30
FCL/PL	0,15	1,02	-1,91	-0,09	0,08	0,36	5,61
FCL/AT	0,04	0,14	-0,49	-0,02	0,04	0,10	0,32
LAJIR/PL	0,42	2,18	-2,40	0,03	0,16	0,28	13,76
LAJIR/AT	0,06	0,08	-0,14	0,03	0,06	0,11	0,22
LAJIDA/PE	0,41	0,68	-0,37	0,15	0,35	0,55	4,19
LAJIDA/AT	0,09	0,08	-0,12	0,03	0,09	0,12	0,25
LL/PL	-0,09	0,49	-2,29	-0,11	0,08	0,17	0,26
LL/AT	0,02	0,08	-0,20	-0,01	0,03	0,06	0,21
LAIR/PL	-0,07	0,52	-2,20	-0,12	0,10	0,19	0,37
DIVLP/PL	2,61	12,50	-0,35	0,18	0,40	0,74	80,54
DIVLP/AT	0,20	0,16	0,00	0,07	0,15	0,31	0,76
NOWC/AT	0,24	0,21	-0,36	0,08	0,19	0,38	0,66
PL/PE	1,43	1,95	-6,70	0,92	1,34	2,01	9,02
PL/AT	0,30	0,47	-2,15	0,23	0,37	0,45	0,86
DIV/PL	3,54	15,93	-0,59	0,42	0,88	1,56	102,82
DIV/AT	0,34	0,21	0,02	0,22	0,31	0,41	1,26
FCL/RLO	50,47	333,34	-31,26	-0,02	0,08	0,20	2132,44
FCL/VA	108,35	685,12	-57,67	-1,13	-0,02	1,23	4386,74
LAJIR/VA	246,62	1581,21	-38,68	-0,60	0,37	1,36	10124,25
LAJIDA/RLO	120,02	768,70	-8,23	0,06	0,13	0,27	4922,03
RB/RLO	0,30	0,21	-0,18	0,18	0,25	0,41	1,00
LL/RLO	127,11	816,35	-13,68	-0,02	0,06	0,11	5226,80
LL/VA	261,93	1679,31	-61,69	-0,24	0,35	1,94	10752,27
VA/IMOB	-0,73	4,09	-24,83	-0,28	-0,08	0,44	3,44
VA/AT	-0,02	0,15	-0,65	-0,06	-0,01	0,05	0,23
VA/RLO	-0,07	0,63	-3,72	-0,13	-0,03	0,21	0,57
CAP/RLO	0,33	0,89	0,01	0,04	0,08	0,21	4,30
AC/RLO	37,02	200,23	0,13	0,38	0,70	1,32	1278,56
EST/RLO	2,07	11,83	0,00	0,06	0,14	0,31	75,94
NOWC/RLO	31,25	176,26	-0,63	0,16	0,42	0,91	1125,37
CAR/RLO	14,78	83,56	0,04	0,12	0,19	0,29	533,34
RLO/AT	0,67	0,67	0,00	0,25	0,48	0,89	3,56