



UNIVERSIDADE FEDERAL DO CEARÁ
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, ATUÁRIA E CONTABILIDADE
PROGRAMA DE ECONOMIA PROFISSIONAL
MESTRADO EM ECONOMIA DE EMPRESAS

RONDINELLY COELHO RODRIGUES

**PREVISÃO DE INSOLVÊNCIA BASEADO EM DADOS CONTÁBEIS DE EMPRESAS
BRASILEIRAS LISTADAS NA BOLSA DE VALORES NOS ANOS DE 2018 E 2019**

FORTALEZA

2021

RONDINELLY COELHO RODRIGUES

PREVISÃO DE INSOLVÊNCIA BASEADO EM DADOS CONTÁBEIS DE EMPRESAS
BRASILEIRAS LISTADAS NA BOLSA DE VALORES NOS ANOS DE 2018 E 2019

Dissertação apresentada ao Programa de Economia Profissional do Centro de Aperfeiçoamento de Economistas do Nordeste (CAEN) da Universidade Federal do Ceará, como requisito parcial para a obtenção do grau de Mestre em Economia de Empresas.

Orientador: Prof. Dr. Sérgio Aquino de Souza

FORTALEZA

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

R616p Rodrigues, Rondinely Coelho.

Previsão de insolvência baseado em dados contábeis de empresas brasileiras listadas na bolsa de valores nos anos de 2018 e 2019 / Rondinely Coelho Rodrigues. – 2021.
36 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Faculdade de Economia, Administração, Atuária e Contabilidade, Mestrado Profissional em Economia de Empresas, Fortaleza, 2021.
Orientação: Prof. Dr. Sérgio Aquino de Souza.

1. Insolvência. 2. Logit. 3. Lasso. 4. Contabilidade. I. Título.

CDD 330

RONDINELLY COELHO RODRIGUES

PREVISÃO DE INSOLVÊNCIA BASEADO EM DADOS CONTÁBEIS DE EMPRESAS
BRASILEIRAS LISTADAS NA BOLSA DE VALORES NOS ANOS DE 2018 E 2019

Dissertação apresentada ao Programa de Economia Profissional do Centro de Aperfeiçoamento de Economistas do Nordeste (CAEN) da Universidade Federal do Ceará, como requisito parcial para a obtenção do grau de Mestre em Economia de Empresas.

Aprovada em: 02 de Junho de 2021

BANCA EXAMINADORA

Prof. Dr. Sérgio Aquino de Souza (Orientador)
Universidade Federal do Ceará (CAEN/UFC)

Prof. Dr. Fabrício Carneiro Linhares
Universidade Federal do Ceará (CAEN/UFC)

Prof. Dr. Rafael Barros Barbosa
Universidade Federal do Ceará (DEA/UFC)

AGRADECIMENTOS

A Deus, pelo dom da vida, Aos meus pais, pelos ensinamentos e amor incondicional, A minha esposa, por tudo que vivemos juntos, A minha filha, por ser minha força diária, Aos meus irmãos, por todo carinho, Aos meus sócios, por todo incentivo e parceria, Ao meu orientador, pela paciência e orientação.

RESUMO

A utilização de modelos para previsão de insolvência e falência de empresas são temas comuns na literatura. É de interesse dos gestores e demais interessados, como acionistas, utilizar indicadores para tentar identificar se uma empresa está próxima a falência ou não. O objetivo desse trabalho é utilizar um modelo econométrico para previsão de insolvência. Utilizou-se o modelo de regressão pós-lasso logit para prever a insolvência de empresas brasileiras listadas na B3 para os anos de 2018 e 2019. Os dados utilizados consistem nos demonstrativos financeiros dessas empresas. Como variáveis explicativas foram utilizados índices contábeis da literatura. O ano de 2019 foi utilizado para classificar entre empresas solventes e insolventes, já o ano de 2018 foi utilizado para estimar os modelos, de modo que haja uma diferença temporal entre os indicadores contábeis e o evento de insolvência. Para efeito comparativo foram estimados outros dois modelos logit com variáveis diferentes. Os resultados mostram bom desempenho preditivo dos modelos, bem como bom ajuste aos dados. O modelo pós-lasso logit foi o de melhor performance preditiva quando comparado aos outros dois modelos, com o percentual corretamente previsto, sensibilidade e especificidade todos acima de 90%.

Palavras-chave: Insolvência. Logit. Lasso. Contabilidade.

ABSTRACT

The use of models to predict company insolvency and bankruptcy are common themes in the literature. It is in the interest of managers and other stakeholders, such as shareholders, to use indicators to try to identify whether a company is close to bankruptcy or not. The objective of this work is to use an econometric model to predict insolvency. The post-lasso logit regression model was used to predict the insolvency of Brazilian companies listed on B3 for the years 2018 and 2019. The data used consist of the financial statements of these companies. As explanatory variables, accounting indices from the literature were used. The year 2019 was used to classify between solvent and insolvent companies, while the year 2018 was used to estimate the models, so that there is a time difference between the accounting indicators and the insolvency event. For comparative purposes, two other logit models with different variables were estimated. The results show good predictive performance of the models, as well as a good fit to the data. The post-lasso logit model had the best predictive performance when compared to the other two models, with the correctly predicted percentage, sensitivity and specificity all above 90%.

Keywords: Insolvency. Logit. Lasso. Accounting.

LISTA DE FIGURAS

Figura 1 – Curva ROC do modelo 1.	29
Figura 2 – Curva ROC do modelo 2.	29
Figura 3 – Curva ROC do modelo 3.	30
Figura 4 – Curvas ROC dos três modelos.	30

LISTA DE TABELAS

Tabela 1 – Variáveis explicativas contábeis a serem testadas no modelo.	22
Tabela 2 – Estatísticas Descritivas das Variáveis Explicativas.	23
Tabela 3 – Resultados do modelo pós-Lasso Logit.	26
Tabela 4 – Resultados do modelo de Guimarães e Filho (2018) com os dados da pesquisa.	27
Tabela 5 – Resultados do modelo pós-Lasso Logit com a inclusão das variáveis LA- CUM.AT e PE.AT.	27
Tabela 6 – Efeitos Marginais do Modelo pós-Lasso Logit	28
Tabela 7 – Efeitos Marginais do modelo de Guimarães e Filho (2018) com os dados da pesquisa.	28
Tabela 8 – Efeitos Marginais do modelo pós-Lasso Logit com a inclusão da variável LACUM.AT	28
Tabela 9 – Percentual Corretamente Previsto do Modelos.	31
Tabela 10 – Sensibilidade dos Modelos.	32
Tabela 11 – Especificidade dos Modelos.	32
Tabela 12 – Taxas de Falsos Positivos dos Modelos.	33
Tabela 13 – Taxas de Falsos Negativos dos Modelos.	33

SUMÁRIO

1	INTRODUÇÃO	10
2	FUNDAMENTAÇÃO TEÓRICA	12
2.1	Insolvência e Falência	12
2.2	Contabilidade	12
2.3	Modelos de Previsão de Insolvência	13
3	FUNDAMENTAÇÃO TEÓRICA	16
3.1	Insolvência e Falência	16
3.2	Contabilidade	16
3.3	Modelos de Previsão de Insolvência	17
4	METODOLOGIA	20
4.1	Base de dados	20
4.2	Proxy de insolvência	21
4.3	Variáveis Explicativas	21
4.4	Modelo Logit	22
4.5	Método Lasso para Seleção de Variáveis	24
5	RESULTADOS	26
6	CONCLUSÕES	34
	REFERÊNCIAS	35

1 INTRODUÇÃO

Muitas empresas não conseguem honrar seus compromissos financeiros, os motivos que levam as mesmas a tal situação podem ser dos mais diversos, é comum que empresários e presidentes de organizações utilizem como principal causa as crises econômicas ou financeiras na maioria dos casos. As dívidas que são acumuladas ao longo de anos chegam a levar ao encerramento das atividades da empresa. Entretanto, sabe-se que mesmo em momento de crise mundial, algumas organizações continuam em crescimento. Levando ao questionamento de que o motivo das falências são as crises econômicas. Portanto, nota-se que são vários os fatores que podem impactar no sucesso ou insucesso de determinada instituição, e a gestão, ou até mesmo, a falta de uma boa gestão se enquadra como um fator determinante para que empresas venham a falir.

Dessa forma, uma das principais ferramentas que auxiliam empresários na tomada de decisão, é a contabilidade, conhecer e saber interpretar às informações trazidas pela contabilidade tem papel essencial para aquele que pretende gerir qualquer empresa. A ciência contábil vem evoluindo bastante ao longo dos anos, o que antes era utilizado apenas para o registro dos fatos que impactavam na mutação do patrimônio da empresa, cada vez mais vem sendo utilizada no auxílio à tomada de decisão. É por meio das informações contábeis que os gestores realmente ficam subsidiados para tomarem as decisões mais assertivas, porém, é também através dessas informações que credores conseguem calcular níveis de endividamento, analisar probabilidades de recebimentos por parte de devedores. Assim, instituições financeiras, fornecedores de matéria-prima, prestadores de serviços e até mesmo os trabalhadores possuem informações, que, se bem analisadas, irão trazer uma maior segurança em suas operações e relações com diversas empresas do Brasil.

No cenário brasileiro há um crescimento no índice de empresas que encerram suas atividades por não conseguirem honrar suas dívidas, seja com bancos, fornecedores ou até mesmo com seus funcionários. De acordo com a recente publicação do IBGE (2019), a taxa de sobrevivência das empresas que foram abertas em 2012, foi de 78,90% após 1 ano de funcionamento, 64,50% após 2 anos de funcionamento, 55% após 3 anos de funcionamento, 47,20% após 4 anos de funcionamento e 39,80% após 5 anos de funcionamento. Portanto, pode-se concluir que aproximadamente 60% das empresas do Brasil fecham suas portas em até 5 anos de existência.

Diante de dados tão alarmantes a questão principal dessa pesquisa reside em identifi-

car quais as principais ferramentas para se analisar a previsão de insolvência de empresas listadas na bolsa de valores B3. De modo que diversos interessados possam utilizar das informações contábeis e modelos de previsão de insolvência afim de obter mais segurança em suas decisões.

A correta utilização de modelos de previsão de insolvência também permite reduzir os custos associados ao processo de falência e diminuem a assimetria de informação entre credores e devedores, administradores e acionistas (GUIMARÃES; FILHO, 2018).

Na literatura de insolvência e falência os modelos mais utilizados são os de análise discriminante e regressão logística. Como variáveis explicativas são utilizados índices contábeis como, por exemplo, nos trabalhos de Kanitz (1978) e Guimarães e Filho (2018).

Neste trabalho utilizou-se da regressão logística com o método lasso para seleção de variáveis. O modelo estimado teve bom desempenho preditivo e bom ajuste aos dados.

Este trabalho está dividido em cinco seções, contando com essa parte introdutória. A segunda seção é referente a fundamentação teórica e revisão de literatura dos modelos de previsão de insolvência. A terceira seção consiste na metodologia empregada nesse trabalho, descrevendo a base de dados e modelos econométricos. A quarta seção apresenta os resultados da metodologia. Por fim, a quinta seção destina-se às conclusões.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção serão abordados os conceitos de insolvência, contabilidade e revisão dos modelos de previsão de insolvência.

2.1 Insolvência e Falência

Para se ter um maior entendimento sobre o assunto a ser estudado, é importante que se tenha total conhecimento que insolvência e falência, apesar de serem termos que estão totalmente ligados e são muitas vezes confundidos, são dois conceitos distintos. A Insolvência é definida como o estado no qual a empresa torna-se impossibilitada de cumprir determinado compromisso. Enquanto que a falência vem a ser definida como “procedimentos que são empreendidos sob leis falimentares quando a corporação está inapta a pagar ou obter acordo com seus credores sem intermediação da justiça” (IUDÍCIBUS; LOPES, 2004). Logo, percebe-se que a condição de insolvência, pode ser um indicador de que a falência de determinada empresa pode estar por vir, vale salientar que esse indicador não deve ser considerado isoladamente afim de se extrair qualquer conclusão a respeito da falência.

Segundo Carter e Auken (2006), as empresas que estão próximo à falência costumam se focar em problemas financeiros, quando, na verdade, estes são, geralmente, apenas sintomas de deficiência operacional muito maior.

É nesse sentido que os gestores e demais interessados nas informações contábeis de empresas não devem se prender apenas a um tipo de indicador para prever se uma empresa está próxima a falência ou não, tanto os fatores econômicos externos, como fatores operacionais internos da instituição, podem ser responsáveis pelo insucesso da mesma. Entretanto, os fatores internos são os de maior atenção e familiaridade entre os gestores, portanto, modelos de previsão de falência devem levar em conta tais fatores.

2.2 Contabilidade

A contabilidade é uma grande aliada das empresas para previsão de insolvência, é através das informações contábeis que os dados são analisados e transformados em indicadores importantíssimos para o auxílio na gestão de qualquer instituição, através das informações contábeis que as tomadas de decisão podem ser tomadas da maneira mais assertiva possível, com informações, claras, precisas e confiáveis. “A informação gerencial contábil é uma das fontes

informacionais primárias para a tomada de decisão e controle nas empresas” (ATKINSON *et al.*, 2000, p. 45).

Para que a contabilidade venha atingir o seu objetivo, que é subsidiar gestores e demais interessados na tomada de decisão, a contabilidade deve possuir um ótimo sistema gerencial de informação contábil. Utilizando informações de várias funções organizacionais diferentes ATKINSON *et al.* (2000).

É através das demonstrações contábeis que as informações são extraídas, entre as principais demonstrações, temos o balanço patrimonial. “O balanço patrimonial é um retrato da empresa, é uma maneira conveniente de organizar e resumir o que a empresa possui e o que a empresa deve, e a diferença entre os dois num dado momento” (ROSS; WESTERFIELD, 2000, p. 58).

Fazendo uma análise das demonstrações contábeis pode-se ter condição de fazer uma avaliação da situação patrimonial da empresa, essa avaliação pode ser tanto em relação ao passado da empresa, através das demonstrações financeiras, dando dessa forma importantes instrumentos que podem ser usados no auxílio à tomada de decisão por parte dos gestores, como também em relação ao futuro, no auxílio às projeções para um futuro empresarial de sucesso.

“Os balanços, principalmente em se tratando de uma análise para avaliação de tendência, devem ser os mais resumidos possível. Se alguma área de problemas for diagnóstica, a análise pode ser aprofundada a qualquer momento” (IUDÍCIBUS, 1980, p. 71).

Com o uso de técnicas contábeis e cruzamento de informações que o profissional consegue transformar os dados de uma demonstração contábil e indicador para tomada de decisão, logo, temos os indicadores de liquidez da empresa, os índices de liquidez medem a capacidade que a empresa tem de liquidar os seus compromissos, essa capacidade de pagamento pode ser tratada no prazo imediato, curto prazo e longo prazo. Para o tratamento desses índices pode-se dizer que quanto maior ele for melhor para a empresa.

2.3 Modelos de Previsão de Insolvência

A previsão de insolvência vem trazer a possibilidade de que os interessados nas informações possam prever a capacidade da empresa de honrar os compromissos, seja de curto prazo, seja de longo prazo. Ao longo dos anos, muitos estudos foram feitos a respeito do assunto, logo, diversos modelos de previsão de insolvência foram desenvolvidos, sempre baseados em informações contábeis, como por exemplo os trabalhos de Fitzpatrick (1932), Breaver (1966) e

Kanitz (1978). Nesse trabalho foi abordado alguns dos estudos da literatura.

Iudícibus (2010) afirma que o primeiro trabalho que se tem notícia sobre o tema é o de Fitzpatrick (1932). De acordo com esse estudo, os índices contábeis mais significativos são Patrimônio Líquido sobre Passivo e Lucro Líquido sobre Patrimônio Líquido, isso foi constatado após fazer um levantamento aleatório em 19 empresas solventes e insolventes, ele acreditou que esses índices seriam os mais importantes no fornecimento de informações relevantes ligados ao risco de inadimplência.

Entre 1954 e 1964, Breaver (1966) estudou dados contábeis de 79 empresas que tinham problemas para honrar seus compromissos e cruzou com esses dados com outras 79 empresas que possuíam uma estrutura financeira saudável, considerando o mesmo ramo de atividades e mesmo volume de ativo, após análise e estudo aprofundando, ele constatou que o melhor indicador financeiro para discriminar empresas insolventes é o fluxo de caixa sobre o total do endividamento.

No cenário brasileiro o principal modelo de previsão de insolvência é o de Kanitz (1978), conhecido como Termômetro de Kanitz. O modelo utilizado pelo autor foi o de análise discriminante, utilizando cinco índices contábeis construídos através dos balanços patrimoniais de empresas brasileiras no período de 1972 e 1974.

Baseado em administração de crédito e previsão de insolvência, Silva (1983), por meio da ferramenta de estatística de análise discriminante, introduzindo novos índices financeiros e testando outros fatores, como segmentação das empresas e horizonte do tempo, demonstrou que os índices que melhor classificam as empresas, variam ano a ano antes da ocorrência da insolvência ou de acordo com o setor de atividade, desenvolvendo assim equações distintas para indústria e comércio.

Na revisão de Pereira e Martins (2015) foram analisados 48 estudos brasileiros de previsão de insolvência para o período de 1974 e 2015. Desses estudos, menos de 25% utilizam o modelo de regressão logística.

Na literatura mais recente temos o trabalho de Guimarães e Filho (2018) que utiliza dados de empresas brasileiras para previsão de falência. O modelo de regressão estimado pelos autores consiste no modelo de regressão logística com seleção de variável pelo método *backward stepwise*. As variáveis explicativas são índices contábeis utilizados em outros estudos da literatura, essas variáveis serão as mesmas utilizadas neste trabalho. Os resultados mostram que apenas dois índices contábeis são significantes para a previsão de falência e o desempenho

preditivo dos modelos estimados foram elevados.

Portanto, percebe-se que existem diversos estudos que trazem modelos de previsão de insolvência de empresas na literatura e no cenário brasileiro. Diversas técnicas estatísticas são aplicadas nos estudos, sendo comum a análise discriminante. Todos os trabalhos são baseados em informações contábeis. Não há uma melhor metodologia estabelecida pela literatura, tendo em vista que há diversos fatores que não são captados pelos modelos e que podem impactar diretamente na capacidade de uma empresa honrar com seus compromissos financeiros. Além disso, diferentes tipos de empresas podem ter diferentes tipos de modelos de previsão, como por exemplo, empresas financeiras e não financeiras.

Conclui-se que, com o desenvolvimento de técnicas estatísticas mais recentes, há espaço na literatura para a construção de novos modelos de previsão de insolvência. No caso desse trabalho serão utilizadas técnicas de aprendizagem estatística (*statistical learning*), também chamado de aprendizagem de máquina (*machine learning*). De acordo com Hastie *et al.* (2017), a aprendizagem estatística refere-se a um vasto conjunto de ferramentas para a compreensão de dados, essas ferramentas podem ser classificadas como supervisionadas ou não supervisionadas. O aprendizado estatístico supervisionado envolve a construção de um modelo estatístico para prever ou estimar uma saída (*output*) com base em uma ou mais entradas (*inputs*). Nesse trabalho será utilizada a técnica de seleção de variáveis chamada de Lasso.

3 FUNDAMENTAÇÃO TEÓRICA

Nesta seção serão abordados os conceitos de insolvência, contabilidade e revisão dos modelos de previsão de insolvência.

3.1 Insolvência e Falência

Para se ter um maior entendimento sobre o assunto a ser estudado, é importante que se tenha total conhecimento que insolvência e falência, apesar de serem termos que estão totalmente ligados e são muitas vezes confundidos, são dois conceitos distintos. A Insolvência é definida como o estado no qual a empresa torna-se impossibilitada de cumprir determinado compromisso. Enquanto que a falência vem a ser definida como “procedimentos que são empreendidos sob leis falimentares quando a corporação está inapta a pagar ou obter acordo com seus credores sem intermediação da justiça” (IUDÍCIBUS; LOPES, 2004). Logo, percebe-se que a condição de insolvência, pode ser um indicador de que a falência de determinada empresa pode estar por vir, vale salientar que esse indicador não deve ser considerado isoladamente afim de se extrair qualquer conclusão a respeito da falência.

Segundo Carter e Auken (2006), as empresas que estão próximo à falência costumam se focar em problemas financeiros, quando, na verdade, estes são, geralmente, apenas sintomas de deficiência operacional muito maior.

É nesse sentido que os gestores e demais interessados nas informações contábeis de empresas não devem se prender apenas a um tipo de indicador para prever se uma empresa está próxima a falência ou não, tanto os fatores econômicos externos, como fatores operacionais internos da instituição, podem ser responsáveis pelo insucesso da mesma. Entretanto, os fatores internos são os de maior atenção e familiaridade entre os gestores, portanto, modelos de previsão de falência devem levar em conta tais fatores.

3.2 Contabilidade

A contabilidade é uma grande aliada das empresas para previsão de insolvência, é através das informações contábeis que os dados são analisados e transformados em indicadores importantíssimos para o auxílio na gestão de qualquer instituição, através das informações contábeis que as tomadas de decisão podem ser tomadas da maneira mais assertiva possível, com informações, claras, precisas e confiáveis. “A informação gerencial contábil é uma das fontes

informacionais primárias para a tomada de decisão e controle nas empresas” (ATKINSON *et al.*, 2000, p. 45).

Para que a contabilidade venha atingir o seu objetivo, que é subsidiar gestores e demais interessados na tomada de decisão, a contabilidade deve possuir um ótimo sistema gerencial de informação contábil. Utilizando informações de várias funções organizacionais diferentes ATKINSON *et al.* (2000).

É através das demonstrações contábeis que as informações são extraídas, entre as principais demonstrações, temos o balanço patrimonial. “O balanço patrimonial é um retrato da empresa, é uma maneira conveniente de organizar e resumir o que a empresa possui e o que a empresa deve, e a diferença entre os dois num dado momento” (ROSS; WESTERFIELD, 2000, p. 58).

Fazendo uma análise das demonstrações contábeis pode-se ter condição de fazer uma avaliação da situação patrimonial da empresa, essa avaliação pode ser tanto em relação ao passado da empresa, através das demonstrações financeiras, dando dessa forma importantes instrumentos que podem ser usados no auxílio à tomada de decisão por parte dos gestores, como também em relação ao futuro, no auxílio às projeções para um futuro empresarial de sucesso.

“Os balanços, principalmente em se tratando de uma análise para avaliação de tendência, devem ser os mais resumidos possível. Se alguma área de problemas for diagnóstica, a análise pode ser aprofundada a qualquer momento” (IUDÍCIBUS, 1980, p. 71).

Com o uso de técnicas contábeis e cruzamento de informações que o profissional consegue transformar os dados de uma demonstração contábil e indicador para tomada de decisão, logo, temos os indicadores de liquidez da empresa, os índices de liquidez medem a capacidade que a empresa tem de liquidar os seus compromissos, essa capacidade de pagamento pode ser tratada no prazo imediato, curto prazo e longo prazo. Para o tratamento desses índices pode-se dizer que quanto maior ele for melhor para a empresa.

3.3 Modelos de Previsão de Insolvência

A previsão de insolvência vem trazer a possibilidade de que os interessados nas informações possam prever a capacidade da empresa de honrar os compromissos, seja de curto prazo, seja de longo prazo. Ao longo dos anos, muitos estudos foram feitos a respeito do assunto, logo, diversos modelos de previsão de insolvência foram desenvolvidos, sempre baseados em informações contábeis, como por exemplo os trabalhos de Fitzpatrick (1932), Breaver (1966) e

Kanitz (1978). Nesse trabalho foi abordado alguns dos estudos da literatura.

Iudícibus (2010) afirma que o primeiro trabalho que se tem notícia sobre o tema é o de Fitzpatrick (1932). De acordo com esse estudo, os índices contábeis mais significativos são Patrimônio Líquido sobre Passivo e Lucro Líquido sobre Patrimônio Líquido, isso foi constatado após fazer um levantamento aleatório em 19 empresas solventes e insolventes, ele acreditou que esses índices seriam os mais importantes no fornecimento de informações relevantes ligados ao risco de inadimplência.

Entre 1954 e 1964, Breaver (1966) estudou dados contábeis de 79 empresas que tinham problemas para honrar seus compromissos e cruzou com esses dados com outras 79 empresas que possuíam uma estrutura financeira saudável, considerando o mesmo ramo de atividades e mesmo volume de ativo, após análise e estudo aprofundando, ele constatou que o melhor indicador financeiro para discriminar empresas insolventes é o fluxo de caixa sobre o total do endividamento.

No cenário brasileiro o principal modelo de previsão de insolvência é o de Kanitz (1978), conhecido como Termômetro de Kanitz. O modelo utilizado pelo autor foi o de análise discriminante, utilizando cinco índices contábeis construídos através dos balanços patrimoniais de empresas brasileiras no período de 1972 e 1974.

Baseado em administração de crédito e previsão de insolvência, Silva (1983), por meio da ferramenta de estatística de análise discriminante, introduzindo novos índices financeiros e testando outros fatores, como segmentação das empresas e horizonte do tempo, demonstrou que os índices que melhor classificam as empresas, variam ano a ano antes da ocorrência da insolvência ou de acordo com o setor de atividade, desenvolvendo assim equações distintas para indústria e comércio.

Na revisão de Pereira e Martins (2015) foram analisados 48 estudos brasileiros de previsão de insolvência para o período de 1974 e 2015. Desses estudos, menos de 25% utilizam o modelo de regressão logística.

Na literatura mais recente temos o trabalho de Guimarães e Filho (2018) que utiliza dados de empresas brasileiras para previsão de falência. O modelo de regressão estimado pelos autores consiste no modelo de regressão logística com seleção de variável pelo método *backward stepwise*. As variáveis explicativas são índices contábeis utilizados em outros estudos da literatura, essas variáveis serão as mesmas utilizadas neste trabalho. Os resultados mostram que apenas dois índices contábeis são significantes para a previsão de falência e o desempenho

preditivo dos modelos estimados foram elevados.

Portanto, percebe-se que existem diversos estudos que trazem modelos de previsão de insolvência de empresas na literatura e no cenário brasileiro. Diversas técnicas estatísticas são aplicadas nos estudos, sendo comum a análise discriminante. Todos os trabalhos são baseados em informações contábeis. Não há uma melhor metodologia estabelecida pela literatura, tendo em vista que há diversos fatores que não são captados pelos modelos e que podem impactar diretamente na capacidade de uma empresa honrar com seus compromissos financeiros. Além disso, diferentes tipos de empresas podem ter diferentes tipos de modelos de previsão, como por exemplo, empresas financeiras e não financeiras.

Conclui-se que, com o desenvolvimento de técnicas estatísticas mais recentes, há espaço na literatura para a construção de novos modelos de previsão de insolvência. No caso desse trabalho serão utilizadas técnicas de aprendizagem estatística (*statistical learning*), também chamado de aprendizagem de máquina (*machine learning*). De acordo com Hastie *et al.* (2017), a aprendizagem estatística refere-se a um vasto conjunto de ferramentas para a compreensão de dados, essas ferramentas podem ser classificadas como supervisionadas ou não supervisionadas. O aprendizado estatístico supervisionado envolve a construção de um modelo estatístico para prever ou estimar uma saída (*output*) com base em uma ou mais entradas (*inputs*). Nesse trabalho será utilizada a técnica de seleção de variáveis chamada de Lasso.

4 METODOLOGIA

Nesta seção serão explanados a base de dados utilizada, classificação das empresas como insolventes, as variáveis explicativas utilizadas e os modelos de regressão utilizados.

4.1 Base de dados

A base de dados desse estudo é composta pela informações contidas nos relatórios financeiros anuais da Comissão de Valores Mobiliários (CVM), disponibilizados por Perlin *et al.* (2018). Os dados consistem em variáveis contábeis das empresas listadas na B3. A amostra contém os anos de 2018 e 2019, sendo o primeiro ano utilizado para a construção das variáveis explicativas, e o segundo ano é utilizado para a construção da variável explicada que categoriza uma empresa como solvente ou insolvente. Foram retiradas da amostra todas as instituições financeiras ou empresas que prestam algum tipo de serviço financeiro como, por exemplo, seguradoras. Tal procedimento foi adotado, pois como explica Guimarães e Filho (2018), tais empresas tem demonstrações muito diferentes das empresas não financeiras.

Da amostra inicial foram retiradas todas as empresas que não tinham dados de algum demonstrativo, como por exemplo balanço patrimonial ativo, ou não tinham informações para o fim do exercício financeiro de 2018 e 2019. Depois foram retiradas empresas que não tinham observações para alguma variável utilizada no modelo econométrico, ou que o valor dessas variáveis eram zero, impossibilitando a construção de índices contábeis que utilizam tal variável como denominador. Além disso, foram retiradas observações que continham receitas totais negativas. Por fim, foram retiradas as observações em que alguma variável explicativa foi considerada como *outlier*, pela classificação dos percentis 2,5% e 97,5%. No total, a amostra final consiste em observações para 311 empresas não financeiras listadas na B3.

A amostra foi dividida aleatoriamente em duas partes: amostra de treino e amostra de teste. A amostra de treino foi utilizada para a estimação dos modelos econométricos, ela consiste em 70% da amostra inicial com um total de 218 observações. Enquanto que a amostra de teste foi utilizada para testar a capacidade preditiva dos modelos estimados, ela consiste em 30% da amostra inicial com um total de 93 observações.

4.2 Proxy de insolvência

A classificação de empresas como solvente ou insolvente foi através do patrimônio líquido negativo. Ou seja, se ao fim do exercício de 2019 uma empresa tem patrimônio líquido negativo ela é considerada como insolvente. Tal classificação de insolvência vai de acordo com Requião (1998) e Guimarães e Alves (2009). Conforme é explicitado por Guimarães e Filho (2018), uma empresa pode se manter operacional mesmo quando as exigibilidades superam seus ativos, uma vez vencidas suas obrigações com fontes financiadoras terceiras, os ativos da empresa não seriam suficientes para fazer frente aos compromissos financeiros, o que caracterizaria insolvência.

Utilizando essa classificação de insolvência, de um total de 311 empresas na amostra, 28 foram classificadas como insolventes em 2019 o que corresponde a aproximadamente 9% do total de empresas.

4.3 Variáveis Explicativas

As variáveis explicativas utilizadas para estimar os modelos econométricos desse trabalho consistem nas variáveis utilizados por Guimarães e Filho (2018). Utilizou-se apenas as variáveis de indicadores contábeis, excluindo as variáveis *dummies* referentes ao faturamento das empresas e as variáveis relacionadas à bolsa de valores.

A tabela 1 contém as 25 variáveis explicativas que serão testadas nos modelos, suas abreviações e o estudo do qual elas foram retiradas. Nos resultados do estudo de Guimarães e Filho (2018), somente as variáveis LACUM.AT e PE.AT foram consideradas estatisticamente significantes.

A tabela 2 contém as estatísticas descritivas das variáveis explicativas. Nota-se que, em geral, as variáveis não tem uma distribuição simétrica.

As variáveis explicativas foram testadas para normalidade multivariada por meio do teste de Henze e Zirkler (1990). O resultado foi de rejeitar a hipótese nula de que as variáveis seguem uma distribuição normal multivariada, ou seja, os dados não podem ser considerados normalmente distribuídos. O método da análise discriminante tem como uma das suas hipóteses a normalidade dos dados. Com a hipótese de normalidade violada e de acordo com Joseph *et al.* (2009), é preferível utilizar modelos como o logit ao invés da análise discriminante. Portanto, nesse trabalho será utilizado o modelo de regressão logit.

Tabela 1 – Variáveis explicativas contábeis a serem testadas no modelo.

Variável	Abreviação	Estudo de referência
Log da Receita Total*	LGRT	Pereira e Martins (2015)
Lucro Líquido/Ativo Total	LL.AT	Bellovary <i>et al.</i> (2007)
Liquidez Corrente	AC.PC	Bellovary <i>et al.</i> (2007)
Capital de Giro/Ativo Total	CG.AT	Sanvicente <i>et al.</i> (1998)
Lucros Acumulados/Ativo Total	LACUM.AT	Bellovary <i>et al.</i> (2007)
EBIT/Ativo Total	EBIT.AT	Bellovary <i>et al.</i> (2007)
Liquidez Seca	AC.EST.PC	Pereira e Martins (2015)
Dívida Total/Ativo Total	DT.AT	Pereira e Martins (2015)
Ativo Circulante/Ativo Total	AC.AT	Bellovary <i>et al.</i> (2007)
Caixa/Ativo Total	CX.AT	Bellovary <i>et al.</i> (2007)
Passivo Exigível/Ativo Total	PE.AT	Bellovary <i>et al.</i> (2007)
Fluxo de Caixa Operacional/Ativo Total	FCO.AT	Sanvicente <i>et al.</i> (1998)
Passivo Circulante/Ativo Total	PC.AT	Pereira e Martins (2015)
Caixa Operacional/Passivo Exigível	FCO.PE	Sanvicente <i>et al.</i> (1998)
Dívida Total/Patrimônio Líquido	DT.PL	Pereira e Martins (2015)
Fluxo de Caixa Operacional/Dívida Total	FCO.DT	Sanvicente <i>et al.</i> (1998)
EBIT/Despesas Financeiras	EBIT.DFIN	Standard & Poor's (1997)
Passivo Exigível/Patrimônio Líquido	PE.PL	Pereira e Martins (2015)
(Ativo Circulante Estoques)/Ativo Total	AC.EST.AT	Pereira e Martins (2015)
Patrimônio Líquido/Ativo Total	PL.AT	Bellovary <i>et al.</i> (2007)
Empréstimos de Longo Prazo/Ativo Total	EMLP.AT	Bellovary <i>et al.</i> (2007)
Caixa/Passivo Circulante	CX.PC	Pereira e Martins (2015)
Caixa Operacional/Passivo Circulante	FCO.PC	Sanvicente <i>et al.</i> (1998)
Capital de Giro/Patrimônio Líquido	CG.PL	Bellovary <i>et al.</i> (2007)
Log do Ativo Total	LGAT	Pereira e Martins (2015)

Nota: * foi utilizado receita total sem o logaritmo.

Fonte: Guimarães e Filho (2018).

4.4 Modelo Logit

O modelo de regressão utilizado para a previsão de insolvência é o modelo de regressão logística, também chamado de modelo logit. Nesse tipo de modelo temos que a variável dependente y_i é binária, no caso desse estudo y_i assume o valor 1 se a empresa é classificada como insolvente e zero caso contrário. De acordo com Wooldridge (2017), nos modelos de resposta binária, o interesse reside, principalmente, na probabilidade de resposta:

$$P(y = 1|X) = P(y = 1|x_1, \dots, x_n)$$

No modelo logit, a probabilidade $P(y = 1|X)$ é modelada da seguinte forma:

$$P(y = 1|X) = G(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) \quad (4.1)$$

Tabela 2 – Estatísticas Descritivas das Variáveis Explicativas.

Variável	Média	Desvio Padrão	Mínimo	1 Quartil	3 Quartil	Máximo
LGRT	1,95	3,78	0,00	0,02	1,83	24,04
LL,AT	0,01	0,12	-0,61	-0,01	0,08	0,20
AC,PC	23,82	225,26	0,00	0,69	2,08	2978,92
CG,AT	-0,03	0,44	-4,24	-0,05	0,11	0,80
LACUM,AT	-0,26	0,82	-7	-0,1	0	0
EBIT,AT	0,05	0,12	-0,51	0,01	0,11	0,73
AC,EST,PC	23,70	225,27	0,00	0,55	1,85	2978,92
DT,AT	0,66	0,65	0,002	0,32	0,78	4,82
AC,AT	0,22	0,19	0,00	0,08	0,34	0,89
CX,AT	0,05	0,08	0,00	0,003	0,06	0,74
PE,AT	0,40	0,49	0,0002	0,15	0,51	4,51
FCO,AT	0,04	0,12	-0,91	-0,01	0,09	0,70
PC,AT	0,25	0,42	0,00	0,07	0,27	4,26
FCO,PE	0,97	17,69	-69,15	-0,02	0,30	297,10
DT,PL	0,62	45,86	-736,03	0,33	2,38	180,92
FCO,DT	-0,01	1,13	-15,17	-0,01	0,17	2,33
EBIT,DFIN	-29,03	340,85	-2808,20	-3,93	-0,24	2359,04
PE,PL	0,06	35,86	-585,15	0,16	1,37	172,31
AC,EST,AT	0,20	0,16	0,00	0,07	0,29	0,82
PL,AT	0,34	0,65	-3,82	0,22	0,68	1,00
EMLP,AT	0,23	0,28	0,00	0,03	0,34	3,17
CX,PC	7,05	73,20	0,00	0,02	0,47	1069,40
FCO,PC	-26,12	396,06	-6904,87	-0,07	0,59	10,59
CG,PL	0,01	7,57	-84,74	-0,07	0,37	85,12
LGAT	21,12	2,02	11,61	20,15	22,47	25,61

Fonte: Elaboração própria.

em que X_i é um vetor de k variáveis explicativas, β é um vetor de $k + 1$ parâmetros e G é a função de distribuição acumulada de uma variável aleatória logística padrão:

$$G(X\beta) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}} \quad (4.2)$$

A estimação dos parâmetros do modelo logit é feita pelo método da máxima verossimilhança. Dado uma amostra aleatória de tamanho n , o estimador de máxima verossimilhança é o vetor de parâmetros $\hat{\beta}_{MV}$ que maximiza a função log-verossimilhança $L(\beta)$, definida por:

$$L(\beta) = \sum_{i=1}^n y_i \ln[G(X_i\beta)] + (1 - y_i) \ln[1 - G(X_i\beta)] \quad (4.3)$$

Com a estimação dos parâmetros do modelo é possível construir o percentual corretamente previsto que é dado pela razão de predições corretas do modelo e o tamanho da amostra. Definindo $\tilde{y}_i = 1$ quando $G(X_i\hat{\beta}) \geq 0,5$ e $\tilde{y}_i = 0$ quando $G(X_i\hat{\beta}) < 0,5$, a predição é

correta quando $\tilde{y}_i = y_i = 1$ e $\tilde{y}_i = y_i = 0$. O valor 0,5 é chamado de ponto de corte e pode assumir diferentes valores de acordo com os critérios de escolha utilizados (WOOLDRIDGE, 2017).

A escolha do ponto de corte, para os modelos estimados, é feita por meio do índice de Youden (1950), definido por:

$$J = \text{Sensibilidade} + \text{Especificidade} - 1 \quad (4.4)$$

O valor do índice de Youden está no intervalo fechado entre 0 e 1. Quanto mais próximo de 1, menores serão os falsos positivos e falsos negativos previstos pelo modelo. Como os falsos positivos e negativos são os erros do tipo 1 e tipo 2, respectivamente, quanto maior o índice de Youden, menores serão os erros. Portanto, quanto maior o índice maior será a capacidade preditiva do modelo estimado.

A sensibilidade de um modelo de classificação binária, como o logit, é dada pela proporção de “positivos” corretamente identificados. Isto é, a razão entre a quantidade de empresas corretamente identificadas como insolventes e a quantidade total de empresas insolventes. A sensibilidade também é chamada de taxa de positivos verdadeiro (*True Positive Rate*).

A especificidade consiste na proporção de “negativos” corretamente identificados. Isto é, a razão entre a quantidade de empresas corretamente identificadas como solventes e a quantidade total de empresas solventes. A especificidade também é chamada de taxa de negativos verdadeiros (*True Negative Rate*).

Geralmente o índice de Youden é utilizado junto com a curva ROC (*Receiver Operating Characteristic*). Para cada ponto de corte determinado, ou para cada ponto da curva ROC, há um valor associado do índice. O ponto de corte ótimo consiste em calcular o valor máximo do índice de Youden ao longo de toda a curva ROC, maximizando a soma da sensibilidade e especificidade do modelo. Graficamente, o valor do índice de Youden consiste na distância vertical entre a curva ROC e a linha diagonal. Assim, o ponto de corte ótimo é obtido quando essa distância é máxima. De uma maneira mais simples, analisando o gráfico da curva ROC, o ponto de corte ótimo consiste nos pontos mais à esquerda e mais para cima.

4.5 Método Lasso para Seleção de Variáveis

Para selecionar dentre as 25 variáveis explicativas utiliza-se do modelo de regressão lasso para modelos logit desenvolvido por Belloni *et al.* (2013). De acordo com Hastie *et al.*

(2017), o método lasso realiza seleção de variáveis a fim de melhorar a precisão da previsão do modelo de regressão no qual ele é aplicado. Ele seleciona um subconjunto reduzido das variáveis explicativas de um determinado modelo.

Considerando um modelo de regressão linear da seguinte forma

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + u_i \quad (4.5)$$

o vetor de coeficientes lasso, $\hat{\beta}_L$, é aquele que minimizam a seguinte quantidade

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^k |\beta_j| \quad (4.6)$$

onde λ é um parâmetro de afinação que é determinado separadamente. No caso desse trabalho, utiliza-se o λ de Belloni *et al.* (2013) que é uma penalidade baseada nos dados. De acordo com a equação 4.6, o estimador lasso é semelhante ao estimador de mínimos quadrados, porém com a adição de um termo de penalidade à soma dos quadrados dos resíduos.

No caso de um modelo logit, o vetor de coeficientes lasso, $\hat{\beta}_L$, é aquele que minimiza

$$-L(\beta) + \frac{\lambda}{n} \sum_{j=1}^k |\beta_j| \quad (4.7)$$

em que $L(\beta)$ é a função log-verossimilhança em 4.3. No caso do modelo logit, o estimador lasso consiste em adicionar um termo de penalidade à função de máxima verossimilhança.

Como é explicitado por Hastie *et al.* (2017), as estimativas da regressão lasso são variantes de acordo com a escala das variáveis explicativas. Portanto, utiliza-se uma padronização das variáveis através do *z-score*.

A estimação dos parâmetros no método lasso, atribui valores não nulos para os coeficientes das variáveis explicativas selecionadas e valores iguais zero para as demais variáveis. Assim, utiliza-se as variáveis explicativas selecionadas pelo método lasso para estimar um novo modelo logit. Esse tipo de procedimento é chamado de regressão pós-lasso (*post-lasso regression*), no caso pós-lasso logit. O estimador pós-lasso é utilizado para aliviar o viés de encolhimento presente no método lasso (BELLONI *et al.*, 2012).

5 RESULTADOS

A tabela 3 contém os resultados da estimação do modelo pós-lasso logit apresentada na seção 4.5, a estimação foi realizada na amostra de treino. Das 25 variáveis explicativas utilizadas no método lasso, apenas 2 delas foram selecionadas: LL.AT e DT.AT. As duas variáveis apresentaram os sinais esperados, negativo e positivo respectivamente, e elas são estatisticamente significantes ao nível de 1%. O modelo indica que há dois fatores importantes na previsão de insolvência, o primeiro é relacionado à lucratividade (LL.AT) e o outro à estrutura de capital (DT.AT)

Tabela 3 – Resultados do modelo pós-Lasso Logit.

Número de variáveis utilizadas	25			
Número de variáveis selecionadas	2			
Método utilizado	Lasso			
λ	27,39			
Coefficientes	Estimativa	Erro padrão	valor z	Valor-p
Intercepto	-4.71***	0.89	-5.30	0.00
LL.AT	-1.02**	0.40	-2.51	0.01
DT.AT	5.67***	1.74	3.26	0.00
AIC	37.35			
BIC	47.51			
Log Likelihood	-15.68			
Deviance	31.35			
Num. obs.	218			
<i>pseudo-R</i> ²	75,58			

Nota: ***, **, * denotam significante ao nível de 1%, 5% e 10%.

Fonte: Elaboração própria.

A tabela 4 contém os resultados da estimação do modelo logit, com os dados dessa pesquisa, utilizando somente as variáveis que foram selecionadas por Guimarães e Filho (2018). Nota-se que a nova estimação do modelo corrobora com os resultados obtidos por Guimarães e Filho (2018). As duas variáveis explicativas LACUM.AT e PE.AT, são estatisticamente significantes ao nível de 1%. Além disso, elas apresentam os mesmos sinais algébricos dos seus parâmetros, negativo e positivo, respectivamente. Assim como o modelo da tabela 3, as duas variáveis explicativas estão relacionadas à lucratividade (LACUM.AT) e à estrutura de capital (PE.AT).

Comparando os resultados dos modelos presentes nas tabelas 3 e 4 temos que o modelo pós-lasso logit é o de melhor performance. De acordo com o critério BIC para seleção

Tabela 4 – Resultados do modelo de Guimarães e Filho (2018) com os dados da pesquisa.

Coeficientes	Estimativa	Erro padrão	valor z	Valor-p
Intercepto	-3.44***	0.49	-6.99	0.00
LACUM.AT	-2.01***	0.49	-4.13	0.00
PE.AT	2.48***	0.74	3.37	0.00
AIC	58.33			
BIC	68.48			
Log Likelihood	-26.16			
Deviance	52.33			
Num. obs.	218			
<i>pseudo-R</i> ²	54,25			

Nota: ***, **, * denotam significante ao nível de 1%, 5% e 10%.

Fonte: Elaboração própria.

de modelos, o modelo mais adequado é o pós-lasso logit, pois apresenta o menor valor BIC. O mesmo vale se for considerado o critério AIC, ou até mesmo o *pseudo-R*². Portanto, há uma melhora no grau de ajustamento do modelo quando fazemos a seleção de variáveis pelo método lasso.

Podemos confirmar esse resultado com a estimação de um modelo que utiliza as duas variáveis selecionadas no modelo lasso e as duas variáveis do modelo de Guimarães e Filho (2018), a tabela 5 contém os resultados da estimação desse modelo.

Tabela 5 – Resultados do modelo pós-Lasso Logit com a inclusão das variáveis LACUM.AT e PE.AT.

Coeficientes	Estimativa	Erro padrão	valor z	Valor-p
Intercepto	-4.86***	1.03	-4.73	0.00
LL.AT	-1.16**	0.52	-2.25	0.02
DT.AT	4.46**	1.84	2.43	0.02
LACUM.AT	-0.43	0.62	-0.69	0.49
PE.AT	1.19	1.04	1.14	0.25
AIC	39.84			
BIC	56.76			
Log Likelihood	-14.92			
Deviance	29.84			
Num. obs.	218			
<i>pseudo-R</i> ²	73,91			

Nota: ***, **, * denotam significante ao nível de 1%, 5% e 10%.

Fonte: Elaboração própria.

O modelo pós-lasso logit, continua sendo o de melhor ajuste pelos critérios BIC, AIC e *pseudo-R*². Além disso, as variáveis LACUM.AT e PE.AT não são estatisticamente significantes ao nível de 10% no novo modelo estimado. Entretanto, como o objetivo dessa

pesquisa é prever insolvência das empresas, o critério de seleção de modelos deve levar em conta a capacidade preditiva dos mesmos. Para isso é necessário comparar a performance dos modelos estimados na base de teste.

As tabelas 6, 7 e 8 contêm o efeito marginal médio das variáveis dos três modelos estimados. O efeito marginal médio é igual a soma dos efeitos marginais em cada observação dividido pelo tamanho da amostra, ou seja, a média dos efeitos marginais.

No caso do modelos da tabelas 6 e 8, nenhum efeito marginal é estatisticamente significativo.

Tabela 6 – Efeitos Marginais do Modelo pós-Lasso Logit

Variável	Efeito Marginal	Erro padrão	Valor t	Valor-p
LL.AT	-0.0190	0.0182	-1.0435	0.2979
DT.AT	0.1058	0.0838	1.2624	0.2082

Fonte: Elaboração própria.

No modelo da tabela 7, os dois efeitos marginais foram estatisticamente significantes ao nível de 5%. Indicando que em média se reduz a probabilidade de uma empresa ser insolvente em 0,06 dado um aumento de uma unidade da variável LACUM.AT. No caso da variável PE.AT, há um aumento, em média, de 0,07 na probabilidade de insolvência.

Tabela 7 – Efeitos Marginais do modelo de Guimarães e Filho (2018) com os dados da pesquisa.

Variável	Efeito Marginal	Erro padrão	Valor t	Valor-p
LACUM.AT	-0.0619**	0.0264	-2.3412	0.0201
PE.AT	0.0763**	0.0305	2.5004	0.0132

Nota: ***, **, * denotam significativa ao nível de 1%, 5% e 10%.

Fonte: Elaboração própria.

Tabela 8 – Efeitos Marginais do modelo pós-Lasso Logit com a inclusão da variável LACUM.AT

Variável	Efeito Marginal	Erro padrão	Valor t	Valor-p
LL.AT	-0.0211	0.0208	-1.0151	0.3112
DT.AT	0.0814	0.0784	1.0392	0.2999
LACUM.AT	-0.0078	0.0153	-0.5070	0.6127
PE.AT	0.0217	0.0254	0.8545	0.3938

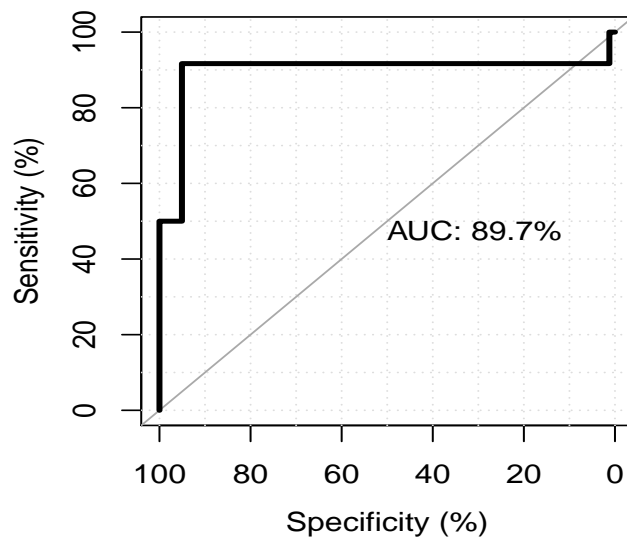
Fonte: Elaboração própria.

Utilizando os três modelos estimados para construir curvas ROC na amostra de teste, pode-se comparar o desempenho preditivo dos modelos em geral. Denota-se por modelo 1 o

modelo pós-lasso logit, por modelo 2 o modelo logit com as variáveis LACUM.AT e PE.AT, e por modelo 3 o modelo com a junção das variáveis dos modelos 1 e 2: LL.AT, DT.AT, LACUM.AT e PE.AT.

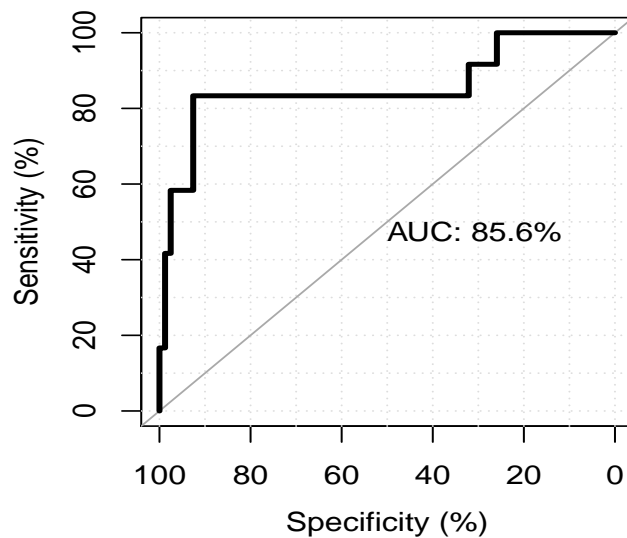
As curvas ROC dos três modelos estão representadas nas figuras 1, 2 e 3. Já a figura 4 apresenta a sobreposição dessas três curvas.

Figura 1 – Curva ROC do modelo 1.



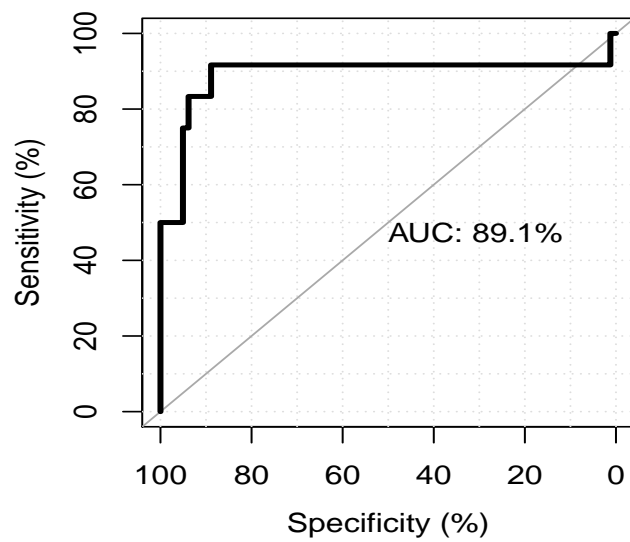
Fonte: Elaboração própria.

Figura 2 – Curva ROC do modelo 2.



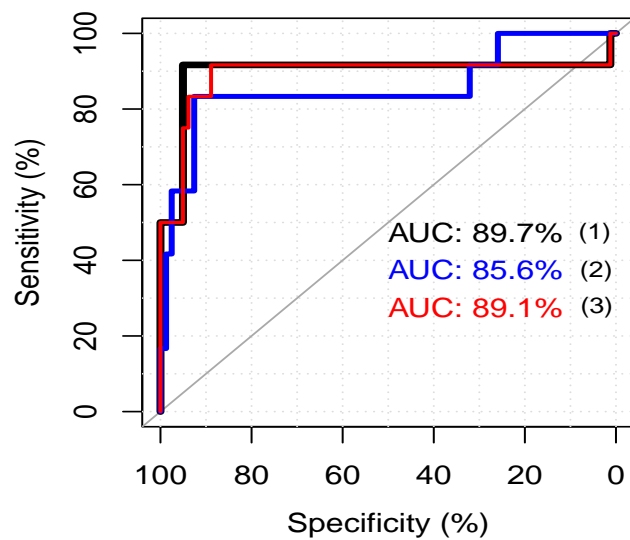
Fonte: Elaboração própria.

Figura 3 – Curva ROC do modelo 3.



Fonte: Elaboração própria.

Figura 4 – Curvas ROC dos três modelos.



Nota: modelo 1 em preto, modelo 2 em azul e modelo 3 em vermelho.

Fonte: Elaboração própria.

Os valores AUC (*Area Under the Curve*) indicam a precisão dos modelos no sentido de separar entre os grupo de empresas solventes e insolventes. O modelo que teve o maior valor AUC foi o modelo 1, seguido pelo modelo 3 e pelo modelo 2. No geral os modelos tiveram boa performance de discriminação, estando no intervalo entre 80% e 90%.

A tabela 9 apresenta o percentual corretamente previsto dos modelos estimados, os valores foram calculados para as amostras de teste e treino. Para cada modelo foi escolhido o ponto de corte que tem o maior valor do índice de Youden (1950), sendo esse calculado na amostra de teste. O mesmo ponto de corte foi utilizado para as previsões na amostra de treino.

Tabela 9 – Percentual Corretamente Previsto do Modelos.

Modelo	Amostra	
	Teste	Treino
Modelo 1 Ponto de Corte: 0,2357	94,62%	98,16%
Modelo 2 Ponto de Corte: 0.1154	91,4%	91,74%
Modelo 3 Ponto de Corte: 0,0640	89,24%	94,04%

Fonte: Elaboração própria.

O modelo 1 foi o que teve o maior percentual corretamente previsto, tanto na amostra de teste quanto na amostra de treino. Em geral todos os modelos tiveram capacidade preditiva elevada, na maioria dos casos maior do que 90%.

A tabela 10 contém as taxas de positivos verdadeiros, ou sensibilidade, dos modelos. Para a amostra de teste os modelos 1 e 3 foram os que performaram melhor, identificando corretamente como insolvente 91,67% das empresas insolventes na amostra de teste. Já na amostra de treino o melhor desempenho foi do modelo 3, classificando corretamente todas as empresas insolventes.

A tabela 11 reporta a taxa de negativos verdadeiros, ou especificidade, dos modelos. O modelo 1 foi o que teve melhor desempenho, tanto na amostra de treino como de teste, classificando corretamente como solvente 95% das empresas solventes na amostra de teste.

A tabela 12 contém as taxas de falsos positivos dos modelos, ou seja, a porcentagem de empresas solventes incorretamente classificadas como insolventes, é o erro do tipo 1. Os

Tabela 10 – Sensibilidade dos Modelos.

Modelo	Amostra	
	Teste	Treino
Modelo 1 Ponto de Corte 0,2357	91,67%	87,5%
Modelo 2 Ponto de Corte 0.1154	83,33%	81,25%
Modelo 3 Ponto de Corte 0,0640	91,67%	100%

Fonte: Elaboração própria.

Tabela 11 – Especificidade dos Modelos.

Modelo	Amostra	
	Teste	Treino
Modelo 1 Ponto de Corte 0,2357	95,06%	99,01%
Modelo 2 Ponto de Corte 0.1154	92,6%	92,57%
Modelo 3 Ponto de Corte 0,0640	88,89%	93,56%

Fonte: Elaboração própria.

modelos 1 e 3 foram os que apresentaram menores valores para a amostra de teste. No caso da amostra de treino o modelo 3 foi o de melhor performance.

A tabela 13 apresenta as taxas de falsos negativos dos modelos, ou seja, a porcentagem de empresas insolventes incorretamente classificadas como solventes, é o erro do tipo 2. Nesse quesito o modelo 1 foi o de menor erro, tanto na base de teste quanto na de treino. Considerando que o melhor modelo é aquele que possui o menor erro de classificar incorretamente uma empresa insolvente como solvente, o modelo 1 teve uma performance bem melhor do que os demais. Sendo menor do que 5% o erro do tipo 2 na amostra de teste e 1% na amostra de treino.

Tabela 12 – Taxas de Falsos Positivos dos Modelos.

Modelo	Amostra	
	Teste	Treino
Modelo 1 Ponto de Corte 0,2357	8,33%	12,5%
Modelo 2 Ponto de Corte 0.1154	16,7%	18,75%
Modelo 3 Ponto de Corte 0,0640	8,33%	0%

Fonte: Elaboração própria.

Tabela 13 – Taxas de Falsos Negativos dos Modelos.

Modelo	Amostra	
	Teste	Treino
Modelo 1 Ponto de Corte 0,2357	4,94%	0,99%
Modelo 2 Ponto de Corte 0.1154	7,41%	7,42%
Modelo 3 Ponto de Corte 0,0640	11,11%	6,43%

Fonte: Elaboração própria.

6 CONCLUSÕES

No modelo pós-lasso logit (modelo 1) as variáveis LL.AT e DT.AT foram ambas estatisticamente significantes ao nível de 1%. Essas duas variáveis levam em conta duas dimensões importantes: lucros líquidos e dívidas totais. Sendo a primeira variável com impacto negativo, na probabilidade de insolvência, e a segunda com impacto positivo. Os efeitos marginais das duas variáveis não foram estatisticamente significantes.

No modelo 2 foi utilizado as variáveis do estudo de Guimarães e Filho (2018), ambas as variáveis foram significantes ao nível de 1%. Entretanto, comparando o grau de ajuste dos modelos 1 e 2, temos que o modelo 1 é melhor ajustado pelos critérios BIC, AIC e *pseudo-R*². O mesmo pode-se concluir do modelo 3.

Em relação ao desempenho preditivo dos modelos, o modelo 1 foi o que teve o melhor desempenho na amostra de teste para todos os critérios: AUC, percentual corretamente previsto, taxas de falso negativo, falso positivo, especificidade e sensibilidade. No caso da sensibilidade e taxa de falso positivo, os modelos 1 e 3 tiveram o mesmo desempenho.

O presente estudo pode ser melhorado com inclusão de outras variáveis para seleção, como variáveis econômicas e outros índices contábeis. Pode-se testar a não linearidade das variáveis, com a inclusão de termos polinomiais, para melhorar a capacidade preditiva. A utilização de uma base de dados em painel com uma amplitude temporal maior pode melhorar as estimativas. Ademais, a metodologia pode ser aplicada para empresas que não estão listadas na B3.

REFERÊNCIAS

- ATKINSON, A.; BANKER, R. D.; KAPLAN, R. S.; YOUNG, M. **Contabilidade gerencial; trad. André Olímpio Nosselman Du Chenoy Castro. São Paulo: Atlas, 2000.**
- BELLONI, A.; CHEN, D.; CHERNOZHUKOV, V.; HANSEN, C. **Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, Wiley Online Library, v. 80, n. 6, p. 2369–2429, 2012.**
- BELLONI, A.; CHERNOZHUKOV, V.; WEI, Y. **Honest confidence regions for a regression parameter in logistic regression with a large number of controls. 2013.**
- BELLOVARY, J. L.; GIACOMINO, D. E.; AKERS, M. D. **A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial education*, JSTOR, p. 1–42, 2007.**
- BREAVER, W. **Financial Ratios as Predicators of Failure, Empirical Research in Accounting: Selected Studies. *Journal of Accounting Research*, v. 5, p. 1–25, 1966.**
- CARTER, R.; AUKEN, H. V. **Small firm bankruptcy. *Journal of Small Business Management*, Wiley Online Library, v. 44, n. 4, p. 493–512, 2006.**
- FITZPATRICK, P. J. **A comparison of the ratios of successful industrial enterprises with those of failed companies. 1932.**
- GUIMARÃES, A. L. d. S.; ALVES, W. O. **Prevendo a insolvência de operadoras de planos de saúde. *Revista de Administração de Empresas*, SciELO Brasil, v. 49, n. 4, p. 459–471, 2009.**
- GUIMARÃES, P. R. F.; FILHO, M. de A. R. **Uma aplicação do modelo de regressão logística na previsão de falência empresarial no Brasil. *Revista Brasileira de Economia de Empresas/Brazilian Journal of Business Economics*, v. 18, n. 2, 2018.**
- HASTIE, T.; JAMES, G. M.; TIBSHIRANI, R.; WITTEN, D. **An introduction to statistical learning: with applications in R. Springer, 2017.**
- HENZE, N.; ZIRKLER, B. **A class of invariant consistent tests for multivariate normality. *Communications in statistics-Theory and Methods*, Taylor & Francis, v. 19, n. 10, p. 3595–3617, 1990.**
- IBGE. **Demografia das empresas e estatísticas de empreendedorismo: 2017. Coordenação de Cadastro e Classificações. Rio de Janeiro, 2019. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv101671.pdf>>. Acesso em: 22 de abril de 2021.**
- IUDÍCIBUS, S.; LOPES, A. B. **Teoria avançada da contabilidade. [S.l: s.n.], 2004.**
- IUDÍCIBUS, S. d. **Análise de balanços 10. ed. São Paulo: Atlas, 2010.**
- IUDÍCIBUS, S. de. **Contabilidade gerencial. Atlas, 1980.**
- JOSEPH, F. H. J.; WILLIAM, C. B.; BARRY, J. B.; ROLPH, E. A. **Multivariate data analysis 7th ed. Upper Saddle River, NJ: Prentice Hall, 2009.**
- KANITZ, S. C. **Como prever falências. São Paulo: McGraw-Hill, 1978.**

- 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*, Universidade Federal de Santa Catarina (UFSC), v. 12, n. 26, p. 163–196, 2015.
- PERLIN, M.; KIRCH, G.; VANCIN, D. **Accessing financial reports and corporate events with GetDFPData.** Available at SSRN 3128252, 2018.
- REQUIÃO, R. **Curso de Direito Falimentar. 2. ed.** Saraiva, São Paulo, 1998.
- ROSS, S. A.; WESTERFIELD, R. W. **Princípios de administração financeira.** Atlas, 2000.
- SANVICENTE, A. Z.; MINARDI, A. *et al.* **Identificação de indicadores contábeis significativos para a previsão de concordata de empresas.** Instituto Brasileiro de Mercado de Capitais, Working Paper, n. 1968, p. 1–12, 1998.
- SILVA, J. P. da. **Administração de crédito e previsão de insolvência.** Atlas, 1983.
- WOOLDRIDGE, J. M. **Introdução à econometria: uma abordagem moderna. 3. ed.** Cengage Learning, São Paulo, 2017.
- YOU DEN, W. J. **Index for rating diagnostic tests.** *Cancer*, Wiley Online Library, v. 3, n. 1, p. 32–35, 1950.