UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DE RIBEIRÃO PRETO
DEPARTAMENTO DE CONTABILIDADE

FLAVIO LUIZ DE MORAES BARBOZA

Previsão de Dificuldades Financeiras em Empresas de Capital Aberto Latino-Americanas

Ribeirão Preto
2022

Prof. Dr. Carlos Gilberto Carlotti Junior
Reitor da Universidade de São Paulo

Prof. Dr. Fabio Augusto Reis Gomes
Diretor da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Prof. Dr. Carlos Alberto Grespan Bonacim
Chefe do Departamento de Contabilidade

FLAVIO LUIZ DE MORAES BARBOZA

Previsão de Dificuldades Financeiras em Empresas de Capital Aberto Latino-Americanas

Tese apresentada para Concurso visando à obtenção do título de Livre-Docente junto ao Departamento de Contabilidade da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo – FEA-RP/USP.

Ribeirão Preto
2022

A minha família.

AGRADECIMENTOS

Agradeço aos meus professores em todos os níveis, em especial àqueles que me orientaram: Elso Drigo Filho, Leonardo Basso, Herbert Kimura e Edward Altman. Aos colegas e parceiros de pesquisa também sou grato, principalmente pelas contribuições e colaborações em nossos estudos.

"Os números governam o mundo."
(Pitágoras de Samos)

RESUMO

BARBOZA, F. L. M. Previsão de Dificuldades Financeiras em Empresas de Capital Aberto Latino-Americanas. 2022. 46 f. Tese (Livre-Docência) – Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, São Paulo, 2022.

A América Latina é uma importante região em desenvolvimento e dinâmica, com um mercado financeiro emergente que tem chamado a atenção de investidores ao redor do mundo. Um aspecto crítico desse mercado é o desempenho das entidades corporativas do setor privado. O mau desempenho e as dificuldades financeiras das empresas são uma preocupação intensa e constante, uma vez que resultam na perda de empregos, perdas para investidores e credores, e a carga resultante sobre a base fiscal e econômica dos países. Este estudo visa examinar modelos de previsão de dificuldades financeiras de empresas de capital aberto em países Latino-Americanos. Usando dados de quase 1.000 empresas nas últimas duas décadas, analisa-se a eficácia das técnicas de Regressão Logística (LR) e *Random Forest* (RF). Vários indicadores financeiros de possíveis dificuldades corporativas são explorados para tentar antever tal evento com horizonte de até três anos. Verificou-se que, em média, o RF supera o LR em termos de poder preditivo e erros tipo I e II. Embora os melhores preditores sejam consistentes ao longo do período de 20 anos para cada modelo, os indicadores variam entre as duas técnicas de modelagem. Adicionalmente, os níveis de desempenho não foram afetados pelo surgimento da pandemia de COVID-19. Em conclusão, esses modelos representam uma ferramenta valiosa para investidores, reguladores, formuladores de políticas governamentais e as próprias empresas na América Latina, auxiliando na tomada de decisões estratégicas e na mitigação de riscos financeiros.

ABSTRACT

BARBOZA, F. L. M. Forecasting Corporate Financial Distress in Latin America. 2022. 46 f. Thesis (Associate Professor) – School of Economics, Business Administration and Accounting at Ribeirão Preto, University of São Paulo, São Paulo, 2022.

Latin America is a significant and dynamic developing region, with an emerging financial market that has attracted the attention of investors worldwide. A critical aspect of this market is the performance of private sector corporate entities. Poor performance and financial difficulties of companies are an intense and constant concern, as they result in job losses, losses for investors and creditors, and the resulting burden on the fiscal and economic base of countries. This study aims to examine corporate's financial distress forecasting models in Latin American countries. Using data from nearly 1,000 companies over the past two decades, the effectiveness of Logistic Regression (LR) and Random Forest (RF) techniques is analyzed. Various financial indicators of potential corporate difficulties are explored to try to anticipate such an event with a horizon of up to three years. It was found that, on average, RF outperforms LR in terms of predictive power and type I and II errors. Although the best predictors are consistent over the 20-year period for each model, the indicators vary between the two modeling techniques. Additionally, performance levels were not affected by the COVID-19 pandemic. In conclusion, these models represent a valuable tool for investors, regulators, government policy-makers, and the companies themselves in Latin America, assisting in strategic decision-making and mitigating financial risks.

LISTA DE ABREVIATURAS

AUC: Área sobre a curva ROC
BS: Brier Score
EBITDA: Earnings Before Interests, Taxes, Depreciation and Amortization
EUA: Estados Unidos da América
FD: Financial Distress
LAJIDA: Lucro Antes de Juros, Impostos, Depreciação e Amortização
PIB: Produto Interno Bruto
ROC: Receiver Operating Characteristic

Lista de Figuras

1 Gráfico de barras mostrando a proporção de casos de FD na América Latina entre 2000 e 2020, considerando quatro definições específicas de FD mostradas na Tabela 2, excluídos os valores faltantes. 29

2 Gráfico de barras comparando a acurácia dos modelos de previsão um ano (primeira barra), dois anos (segunda) e três anos à frente a divulgação dos resultados das empresas com base nos métodos Regressão Logística (LR) e Random Forest (RF) para cada um dos as formas de variáveis dependentes (FD1 a FD4). . 32

3 Brier Score para cada modelo, considerando o horizonte de previsão e o tipo de dificuldade (FD1–FD4). 35

4 Curvas ROC (Falso Positivo x Verdadeiro Positivo). As colunas representam o horizonte de previsão para FD (ou seja, t+1, t+2 e t+3, respectivamente). Cada linha é fornece o tipo de FD (FD1 até FD4). 36

5 Mapa de calor que reflete a Importância de cada variável explicativa para ambas as técnicas de modelagem e todos os horizontes de previsão. Preto significa um melhor preditor (ou determinante) e branco pode ser considerado irrelevante. 38

Lista de Tabelas

1 Total de empresas avaliadas de 2000 a 2020 e capitalização de mercado de cada país em 2019. * Fonte: The World Bank (2021) 23

2 Definições de FDs usadas para realizar este trabalho e o número de casos identificados no banco de dados. 25

3 Variáveis Independentes utilizadas como fonte de informação para os modelos previsores e suas respectivas definições. 26

4 Representação da Matrix de Confusão. Estrutura básica que resume todos os resultados das previsões em uma única tabela. . 29

5 Estatística Descritiva das Observações Empregadas no Estudo. Neste caso, considera-se as empresas em dificuldades como FD1 (ou seja, pelos fundamentos das empresas). 31

6 Erros Tipo I e II dos Modelos Preditivos baseados em LR e RF para cada Horizonte de Previsão. O erro do tipo I mostra a proporção de empresas que solventes de fato, mas o modelo classificou como FD, enquanto o erro do tipo II mostra a porcentagem de empresas FD classificadas como solventes. 34

7 Desempenho de RF com relação a acurácia antes e depois do início da pandemia. Nível de Significância: (*)10%, (**)5%, e (***)1% 40

8 Desempenho de RF com relação aos erros antes e depois do início da pandemia. Nível de Significância: (*)10%, (**)5%, e (***)1% 40

Sumário

1 Introdução **13**

1.1 Objetivos 15

1.2 Contribuições e Implicações 16

2 Referencial Teórico **18**

2.1 Conceitos e Definições 18

2.2 Literatura Relacionada 19

2.3 Pesquisas na América Latina e ao redor do mundo 21

3 Delineamento de Pesquisa **23**

3.1 Dados 23

3.2 Variáveis Dependente e Independentes 24

3.3 Procedimento e Modelagem de Previsão 27

3.4 Método de Análise 29

4 Resultados e Análises **31**

4.1 Previsões ao longo do tempo 32

4.2 Classificação de Importância das Variáveis Independentes 37

4.3 Análise de persistência no tempo e o efeito da pandemia 39

5 Considerações Finais **42**

Referências **43**

1 Introdução

A análise do desempenho financeiro é uma das atividades cotidianas relacionadas à Contabilidade Financeira e tem sua devida utilidade no processo de tomada de decisão.

No âmbito das empresas, que precisam de recursos para financiar suas atividades, a captação está intimamente ligada com os investimentos. Isso porque os administradores financeiros avaliam a necessidade de financiamentos caso tenham projetos rentáveis, ou melhor, que sejam capazes de gerar mais lucro do que o custo de capital, seja ele de terceiros ou próprio.

Do outro lado, há gestores de fundos e investidores locais preocupados com o valor dessas corporações e ainda com a sua qualidade de crédito, ou seja, na capacidade dessas empresas arcar com seus compromissos financeiros e, com isso, terem resultados mais previsíveis, de forma que mostrem seus resultados de forma mais próxima àquilo que consta em seus planos de investimentos. Isso levanta a discussão sobre o que é incerto e o que é risco.

A incerteza está ligada a situações em que os resultados futuros são desconhecidos ou difíceis de serem previstos com base em informações disponíveis. Isso envolve eventos imprevisíveis e não é possível fazer estimativas para diferentes resultados. A incerteza geralmente surge de eventos não esperados, como mudanças repentinas nas condições econômicas, políticas, regulatórias ou outros eventos imprevistos que afetam os mercados financeiros. A incerteza torna mais difícil a tomada de decisões informadas e pode aumentar o receio de tomar decisões em cenários desconhecidos por parte dos investidores.

Já o risco refere-se a situações em que os possíveis resultados de uma decisão ou investimento são conhecidos, pelo menos em termos probabilísticos. Nesse contexto, é possível estimar a probabilidade de ocorrência de diferentes resultados e avaliar suas potenciais consequências. O risco está relacionado à variabilidade dos retornos esperados e é mensurável por meio de alguma ferramenta específica e bem definida, isto é, que se saiba objetivamente como mensurar. Os investidores podem utilizar técnicas como a análise de variância, desvio padrão e *Value-at-Risk* (VaR) para avaliar e gerenciar o risco. Em geral, quanto maior o risco associado a um investimento, maior a possibilidade de obter retornos mais elevados, mas também maiores chances de perdas.

Essa perspectiva pode ser observada sob a ótica da gestão de riscos, em especial dos riscos financeiros. As maiores preocupações giram em torno do risco de mercado e do risco de crédito. O primeiro refere-se ao grau de incerteza quanto ao valor das ações, o que está mais relacionado à expectativa dos inves-

tidores quanto ao valor de uma empresa. O segundo está ligado ao potencial de uma empresa não cumprir com seus compromissos, ocasionando prejuízos aos credores. Em geral, eventos de perdas – como inadimplência ou estresse financeiro – são notadamente importantes e possivelmente mais graves.

Em particular, o estresse financeiro, mais comumente chamado de *Financial Distress* (FD), é um assunto amplamente discutido na literatura (Bali, Del Viva, Lambertides, & Trigeorgis, 2020; Chen, 2011; Khemakhem & Boujelbene, 2018) e de várias formas (Lee & Yeh, 2004; Mousavi & Ouenniche, 2018), principalmente após a crise financeira global de 2008 (Carvalho, Ferreira, & Matos, 2015; Inekwe, Jin, & Valenzuela, 2018; Pindado, Rodrigues, & de la Torre, 2008).

O FD é geralmente estudado em nível corporativo e tem sido aplicado em empresas financeiras e não financeiras. Por exemplo, Avino, Conlon, & Cotter (2019) investigaram bancos e o impacto do *spread* de CDS na possibilidade de falência deles, enquanto Oz & Yelkenci (2017) examinaram empresas não financeiras de vários países na tentativa de desenvolver um modelo generalizado de antecipação de FD.

Compreender e prever esse fenômeno é uma tarefa relevante e impactante para diversos profissionais e investidores (Chen, 2011; Halteh, Kumar, & Gepp, 2018; Khemakhem & Boujelbene, 2018; Masulis & Simsir, 2018; Mousavi & Ouenniche, 2018). Segundo Altman (2018), os investidores precisam ter uma perspectiva de negócio para identificar oportunidades de retorno mais atrativas, os analistas necessitam de ferramentas mais confiáveis e orientar seus clientes com mais clareza, os reguladores são obrigados a monitorar o mercado para evitar crises e contágios, os bancos buscam melhor compreensão da qualidade do crédito, auditores buscando formas mais precisas de encontrar falhas nos dados financeiros que camuflam informações negativas das empresas, consultores exigem modelos que identifiquem pontos estratégicos a serem aprimorados nos negócios. Em suma, os profissionais têm inúmeros motivos e a academia pode ser uma fonte consolidada para atender a essas necessidades.

As origens das empresas de FD incorporam o mundo como um todo. Estudos empíricos relevantes podem ser vistos nos EUA (Acharya, Bharath, & Srinivasan, 2007; Barboza, Kimura, & Altman, 2017; Frydman, Altman, & Kao, 1985; Pindado et al., 2008; Sayari & Mugan, 2017, entre outros), na zona do euro (Altman, Iwanicz-Drozdowska, Laitinen, & Suvas, 2017; Gupta, Gregoriou, & Ebrahimi, 2018; Keasey, Pindado, & Rodrigues, 2015; Klepac & Hampel, 2017; Laitinen & Suvas, 2016), países asiáticos (Altman, Eom, &

Kim, 1995; Alifiah, 2014; Chen, 2011; Geng, Bose, & Chen, 2015) e até em alguns países da América Latina (Altman, Baidya, & Dias, 1979; Altman et al., 2017; Rezende, Montezano, Oliveira, & Lameira, 2017; Wanke, Barros, & Faria, 2015).

Para se ter uma ideia, o mercado de derivativos brasileiro é um dos 3 maiores do mundo em 2021 (FIA, 2021) e representa 7,5% da capitalização das empresas listadas entre os Mercados Emergentes em 2018 (The World Bank, 2021). Nesse mesmo ano, o México foi classificado como a 15^a maior economia mundial, conforme relatório do Banco Mundial (2021). O PIB chileno cresceu consistentemente e apenas diminuiu em circunstâncias críticas (crises financeiras de 2007-08 e 2020 – o primeiro ano da pandemia). Juntos, esses países representam mais de 80% do valor de mercado das empresas de capital aberto na América Latina (The World Bank, 2021).

Apesar dessa visibilidade, essa região apresenta um dos maiores níveis de risco nas bolsas de valores do mundo. Isso ocorre devido à alta volatilidade (Cardona, Gutiérrez, & Agudelo, 2017) e às fragilidades das empresas, o que é evidenciado em seus desempenhos e baixo grau de classificação de crédito.

Embora exista um menor empenho acadêmico na América Latina, talvez pela falta de dados disponíveis e ou investimentos, o ambiente latino-americano tem demonstrado crescente interesse de investidores de todo o mundo. Não obstante, este mercado emergente possui uma dimensão significativa (equivalente ao mercado alemão) e, em contrapartida, é conhecido pela baixa liquidez (Escobari, Garcia, & Mellado, 2017).

1.1 Objetivos

Diante do exposto, este estudo tem como objetivo examinar em profundidade a possibilidade de ocorrência de FD, bem como identificar os principais determinantes (indicadores contábil-financeiros) dessas dificuldades financeiras em empresas latino-americanas.

O propósito é investigar empresas com dados financeiros disponíveis entre 2000 e 2020, focando na previsão de quais empresas enfrentarão dificuldades financeiras ao longo de um, dois e três anos após a divulgação de suas demonstrações financeiras.

Para tanto, duas técnicas são avaliadas: a regressão logística (LR), que é tradicional nesse tipo de pesquisa e utilizada aqui como *benchmark* (Chen, 2011; Gepp & Kumar, 2015; Laitinen & Suvas, 2016); e *Random Forest* (RF), que é um método que demonstrou ter melhor capacidade preditiva em estudos

recentes (e.g., Barboza et al., 2017; Bubna, Das, & Prabhala, 2020; Halteh et al., 2018; Koutanaei, Sajedi, & Khanbabaei, 2015) e sugerido por Gepp & Kumar (2015).

Além disso, fenômenos financeiros são naturalmente mais complexos, apresentando aspectos não lineares de comportamento (Li, Lu, & Srinivasan, 2019), principalmente na América Latina (Romero-Barrutieta, Bulíř, & Rodríguez-Delgado, 2015), o que diminui a qualidade de técnicas como LR (Khemakhem & Boujelbene, 2018).

Em seguida, realiza-se uma análise sobre os indicadores que melhor explicam o FD, o que é possível com ambas as técnicas, embora o RF apenas indique a hierarquia de importância, sem indicar a direção (ou seja, efeito negativo ou positivo).

Uma análise complementar também é feita no sentido de identificar a robustez dos modelos preditivos ao longo dos anos, o que traz a tona a questão da obsolescência dessas ferramentas. Aproveitando-se disso, este estudo verifica se o acontecimento de um evento extremo (no caso, a pandemia da COVID-19) implica em alguma mudança nas previsões.

1.2 Contribuições e Implicações

Os achados mostram que o poder preditivo da modelagem de RF é superior ao LR na maioria dos casos, considerando cinco medidas diferentes de qualidade de desempenho. Mais especificamente, o RF alcança o melhor desempenho ao prever FD em 1 ano antes, confirmado pela área sob a curva ROC (AUC) e Brier Score (BS).

Em termos de acurácia média, o modelo LR é melhor que o RF no mesmo cenário, mas sua taxa de erro do tipo II foi maior neste caso, o que vai ao encontro dos delineamentos de Rezende et al. (2017).

Outra saída reveladora diz respeito à importância das variáveis explicativas para cada modelo que auxiliam na previsão do FD. Os modelos LR sugeriram (em geral) as variáveis País, Setor, Mudança no *price-to-book* e as variáveis do modelo de de Altman (1968), enquanto os modelos RF indicam quase o mesmo, exceto País e Setor, e dando mais relevância aos indicadores relacionados aos lucros.

Assim, algumas implicações podem ser destacadas. Em primeiro lugar, a técnica de *Random Forest* se mostra mais eficiente na previsão de FD e, assim, confirma mais uma vez sua aplicabilidade em estudos ligados a temas de finanças, especialmente para previsões baseadas em indicadores contábil-

financeiros extraídos das demonstrações financeiras corporativas.

Segundo, as empresas latino-americanas representam um mercado crescente em relação ao resto do mundo, destacando a necessidade de mais pesquisas envolvendo essa região. Nesse sentido, este estudo marca um passo importante na evolução deste mercado emergente (Escobari et al., 2017).

Por fim, indicadores relacionados a mercado, rentabilidade e liquidez são relevantes e, além disso, na forma de variabilidade (ou dinâmica) em alguns casos, reforçando os resultados de Barboza et al. (2017); Gupta et al. (2018); Kim & Upneja (2014); Laitinen & Suvas (2016); Mousavi & Ouenniche (2018).

Esta pesquisa empírica contribui tanto para acadêmicos quanto para praticantes, pois apresenta avanços na previsão de FD com horizonte de tempo adequado como também aponta indicadores que sinalizam tal evento.

Com isso, os gestores podem se dedicar de forma objetiva aos problemas enfrentados dentro das empresas, especialmente na América Latina. A análise dessas empresas enriquece a discussão da academia sobre os fatores que levam as corporações a experimentar qualquer preocupação e até mesmo ao fracasso. Assim, a contribuição deste estudo se baseia nos resultados diretos que podem auxiliar claramente uma melhor interpretação do mundo real.

Em termos científicos, este estudo também mostra alguns aspectos empíricos importantes. Primeiramente, discute o desempenho das empresas nos países da América Latina. Em segundo lugar, apresenta uma ferramenta potencial para analistas e investidores melhorarem suas expectativas em relação a este mercado. Terceiro, promove uma discussão sobre indicadores que possam sugerir melhores alternativas para os gestores evitarem problemas financeiros e falência das organizações, na pior das hipóteses.

O restante deste trabalho está organizado da seguinte forma: a seção 2 apresenta resultados proeminentes da área de estudo, ou seja, a literatura relacionada ao tema. A metodologia é descrita juntamente com os dados na seção 3. Os resultados são mostrados na seção 4, seguidos de suas discussões. Por fim, a seção 5 expõe as considerações finais, elencando as conclusões desta tese e propostas de estudos futuros.

2 Referencial Teórico

Empresas em estresse financeiro são examinadas em várias perspectivas diferentes na literatura. Com foco na previsão de FD, é possível encontrar vários artigos (Avino et al., 2019; Chen, 2008; Geng et al., 2015; Halteh et al., 2018; Kim & Upneja, 2014; Mousavi & Ouenniche, 2018; Pindado et al., 2008, entre outros), incluindo os seminais (por exemplo, Altman, 1968).

A seguir, apresenta-se uma série de conceitos e estudos que fundamentam esta pesquisa. Importante destacar que o marco teórico em sua essência ainda é inexistente nesse ramo, o que significa que a base científica consiste na evolução de pesquisas científicas que circundam principalmente as áreas de: Contabilidade, no que concerne as análises de demonstrações financeiras; Finanças, quando se trata da eficiência financeira das empresas e gestão de riscos por investidores e credores, e também pelos próprios gestores; e, Economia, ao se preocupar com o ambiente macro e com a rentabilidade dessas firmas.

2.1 Conceitos e Definições

Basicamente, os estudos científicos nesse campo de pesquisa consideram um evento específico como episódio de dificuldade financeira, como inadimplência e falências, mas o registro deles é bastante difícil de encontrar. Para evitar esse problema, Pindado et al. (2008) sugeriram uma definição de FD corporativa baseada em demonstrações financeiras e valor de mercado. Segundo os autores, uma empresa pode ser considerada em dificuldades financeiras quando duas condições são satisfeitas: primeiro, "seu lucro antes de juros e depreciação e amortização de impostos (LAJIDA, também conhecido pelo termo em inglês, EBITDA) é menor que suas despesas financeiras por dois anos consecutivos"; e, segundo, "uma queda em seu valor de mercado ocorre entre dois períodos consecutivos". Além disso, justificam que o FD tem natureza não estacionária e deve ser analisado levando-se em consideração essa propriedade.

Essa definição foi utilizada por outros pesquisadores e adaptada em algumas circunstâncias. Particularmente, Manzaneque, Priego, & Merino (2016) examinaram empresas espanholas e aceitou apenas uma dessas condições para classificar como FD. Assim, aqui inclui-se mais de uma definição para avaliar as diferentes possibilidades de predição de estresse.

É comum a literatura abordar o termo *Financial Distress* mas, na verdade, tratar de eventos de falência. Isso pode ser visto em diversos trabalhos, tais como, Altman et al. (2017); Liang, Lu, Tsai, & Shih (2016); Frydman et al.

(1985); Altman (1968).

2.2 Literatura Relacionada

Beaver (1966), Altman (1968) e Ohlson (1980) podem ser considerados os artigos fundamentais da pesquisa aqui apresentada.

Beaver (1966) fez uma análise univariada para detectar falência. Seu estudo é considerado o primeiro a propor o modelo univariado, de modo a entender quais indicadores financeiros, um a um, eram diferentes entre empresas falidas e não falidas. Assim, usou amostras pareadas (para empresas do mesmo setor e de tamanho semelhante), que compreendia uma amostra de 79 empresas falidas e 79 empresas não falidas, captadas na base de dados da Moody's. Esses dados abrangiam um período de aproximadamente 10 anos (de 1954 a 1964) e excluía empresas não industriais (por exemplo, serviços públicos, empresas de transporte e instituições financeiras).

Ao avaliar 30 indicadores com base em seu uso em estudos anteriores, os resultados revelaram que houve diferenças significativas nas proporções entre empresas falidas e não falidas durante o período de observação. Destaca-se que as empresas falidas tinham índices de fluxo de caixa em relação à dívida total e rendimento líquido em relação aos ativos totais mais baixos do que as empresas não falidas. Entretanto o modelo de análise é considerado limitado em sua análise, principalmente por não trazer um modelo

Já Altman construiu o conhecido modelo Z-score, que vem sendo aplicado por inúmeros pesquisadores, analistas, investidores, entre outros ao redor do mundo (Altman, 1968).

Assim como Beaver, Altman também foi pioneiro, porém seus avanços foram além do anterior, quando partiu para a abordagem multivariada no estudo de falências. Altman comparou uma amostra de 33 empresas falidas e 33 não falidas nos EUA durante o período de 1946-1965. As empresas estavam no mesmo setor (manufatura) e tinham ativos de tamanho semelhante. Inicialmente, o pesquisador testou um total de 22 indicadores financeiros considerando o que a literatura prévia, incluindo o estudo de Beaver (1966).

Assim, Altman desenvolveu o modelo Z score tendo como ferramenta básica a técnica de Análise Discriminante Múltipla. A principal vantagem dessa técnica é a possibilidade de reduzir uma visão multidimensional em um único escore Halteh et al. (2018). Em suma, o modelo Z-score é dado por:

$$Z = 0,012.X1 + 0,014.X2 + 0,033.X3 + 0,006.X4 + 0,010.X5, \quad (1)$$

tal que $X1$ é a razão Capital de Giro/Total de Ativos, $X2$ é dado por Lucros Retidos/Total de Ativos, $X3$ é Lucros antes Juros e impostos/Ativo total, $X4$ = Valor de mercado do patrimônio líquido/Valor contábil do passivo total e $X5$ = Vendas/Ativo total. Z representa a pontuação da empresa.

Como o escore não foi padronizado, Altman apresentou os pontos de corte para escore. A empresa com escore acima de 2,7 é considerada saudável, ou seja, tem baixa possibilidade de ir à falência. A conclusão é se o escore for menor ou igual a 1,8 mas se estiver entre esses valores, o autor definiu a região cinzenta do seu modelo, o que significa que a empresa nessa faixa de escore deve ficar sob observação e, portanto, sua análise é inconclusiva.

Os resultados apresentados no estudo mostram uma capacidade preditiva expressiva. A precisão do modelo para previsões um ano à frente chega a 95%, mas diminui para 72% quando é de dois ou mais anos antes da falência.

Em um avanço proeminente nos estudos de finanças, Ohlson (1980) implementou a regressão logística neste contexto, que é uma das técnicas estatísticas que requer menos suposições restritivas do que a análise discriminante.

Ohlson (1980) aplicou uma estimativa probabilística de falência em seu estudo e desenvolveu o chamado modelo logit de empresas americanas (105 empresas falidas e 2.058 não falidas) entre os anos 1970 e 1976. Como resultado, o modelo logit de Ohlson apresentava 9 indicadores como variáveis explicativas.

O método é amplamente utilizado em estudos, como Avino et al. (2019); Chen (2008); Klepac & Hampel (2017); Pindado et al. (2008), entre outros.

Frydman et al. (1985) pode ser considerado um dos primeiros trabalhos a investigar FD usando método computacional, sugerindo que esta modelagem é convincente e pode ser útil no processo de tomada de decisão. No caso, os autores dedicaram-se a avaliar a performance de um método denominado "Algoritmo de Reparticionamento Recursivo" e foi um dos primeiros estudos a enfatizar a melhoria na qualidade das previsões ao se aplicar uma técnica não paramétrica. Ainda, apontaram que seus achados podem ser usados por praticantes como uma ferramenta interessante para auxiliar na tomada de decisões.

Chen (2011) trabalhou com mais de 30 medidas financeiras e não financeiras na previsão de empresas em FD. Os resultados mostraram melhor desempe-

nho de modelagem computacional em relação à regressão logística em curtos períodos de previsão, embora esta última tenha alcançado níveis satisfatórios (mais de 90% de precisão em alguns casos). Por outro lado, em uma distância maior do evento, o LR superou o modelo de árvores de decisão.

2.3 Pesquisas na América Latina e ao redor do mundo

Diferentes países e regiões podem ser examinados em termos de FD. Diversos estudos examinaram empresas americanas, como Acharya et al. (2007); Barboza et al. (2017); Frydman et al. (1985); Pindado et al. (2008); Sayari & Mugan (2017). A Europa também é normalmente explorada (e.g., Altman et al., 2017; Gupta et al., 2018; Keasey et al., 2015; Klepac & Hampel, 2017; Laitinen & Suvas, 2016). Outros locais, como países asiáticos, podem ser encontrados com certo interesse dos pesquisadores (Alifiah, 2014; Altman et al., 1995; Chen, 2011; Geng et al., 2015).

Na América Latina, alguns estudos podem ser encontrados. Cardona et al. (2017), por exemplo, investigaram o mercado latino-americano e sua relação com o mercado norte-americano. Neste estudo, os autores destacam o crescimento do mercado latino-americano e sua importância em relação à economia mundial. Além disso, apontam para as instabilidades que esses países vivem continuamente nos últimos anos, bem como o aumento significativo, quase 15 vezes, do investimento estrangeiro em empresas da região.

Rezende et al. (2017) é outro estudo nesta região, mas examinou FD apenas no Brasil. Os autores aplicaram mais de 30 variáveis, com base nas demonstrações financeiras das empresas, dados macroeconômicos e medidas esperadas, mas o crescimento ou mudanças nos índices financeiros não foram testados. O poder de predição foi dado pela AUC (0,82), especificidade (91,3%) e sensibilidade (50%).

O setor financeiro é outro campo de estudo e que precisa ser pesquisado separadamente devido a suas demonstrações seguirem um padrão específico. Por exemplo, Halteh et al. (2018) pesquisou bancos islâmicos usando 18 medidas financeiras, incluindo o Z-score de Altman para detectar o estresse dessas organizações. Os autores aplicaram métodos computacionais, incluindo RF, e inferiram que essa técnica funciona "muito bem".

Gepp & Kumar (2015) compararam o desempenho de LR com a técnica de Árvores de Decisão usando 27 medidas financeiras tendo LR uma acurácia menor. Em particular, os autores sugerem o uso de *Random Forest* como uma técnica potencial a ser empregada.

Na China, Geng et al. (2015) estudaram FD usando mais de 30 indicadores financeiros, considerando um horizonte de 3, 4 e 5 anos de previsão. Eles também avaliaram a importância das variáveis, observando que o EPS é uma das principais medidas para classificar as empresas em FD.

Chen (2008) examinou o efeito da escala de tempo das medidas de governança corporativa na previsão de FD para um, dois e três anos antes do evento usando LR com dados de Taiwan. Esses horizontes de tempo são úteis para prever o FD porque fornecem uma medida de risco interessante para os investidores. Assim, também adota-se esses horizontes neste estudo.

Outros estudos também reforçam a superioridade de métodos baseados em árvores de decisão sobre a regressão logística para prever FD. Por exemplo, Gepp & Kumar (2015) propuseram um modelo de previsão de FD e testaram a qualidade dos modelos usando taxas de erros de custo, como Erro Tipo I, Erro Tipo II, mas não recomendam usar a medida de AUC. Os autores apontaram que os níveis de erro do LR são geralmente maiores do que as árvores de decisão. Novamente, explora-se essa conclusão e inclui-se aqui os erros de tipo I e II como medidas de avaliação do desempenho preditivo.

Quando o foco é a previsão de FD, os resultados mais importantes estão associados ao erro tipo II, ou seja, para classificar as empresas que atingiram o FD, mas foram classificadas como solventes (Altman, 1968; Chen, 2011; Gepp & Kumar, 2015), uma vez que esses erros podem causar perdas financeiras enquanto o erro do tipo I apenas leva a evitar uma oportunidade lucrativa.

3 Delineamento de Pesquisa

3.1 Dados

Os dados para este estudo contempla uma amostra latino-americana que foi extraída da base de dados Thomson Datastream com base em quase 21 anos de dados (2000 a 2020). Ao analisar previsões de dificuldades financeiras corporativas nesta região, é importante considerar um período de tempo abrangente para obter uma visão mais completa e contextualizada.

Esse banco de dados inclui números extraídos de demonstrações financeiras (especialmente, balanço patrimonial e demonstração de resultados) de 808 empresas em 6 países (Argentina, Brasil, Chile, Colômbia, México e Peru), resultando em 10.118 observações (empresa-ano) que contemplam os indicadores financeiros, descritos a seguir. A Tabela 1 mostra o número de empresas estudadas por meio desta base de dados, organizadas por país.

Tabela 1: Total de empresas avaliadas de 2000 a 2020 e capitalização de mercado de cada país em 2019. * Fonte: The World Bank (2021)

País	Número de Empresas	Amostras	Valor de Mercado (USD bi)* (2019)
Argentina	82	1.058	39,4
Brasil	310	3.480	1.187,4
Chile	138	1.954	203,8
Colômbia	48	589	132,0
México	126	1.645	413,6
Peru	104	1.392	98,9
Total	808	10.118	2.075,1

Pela perspectiva da abrangência temporal, a pesquisa cobre duas décadas, permitindo uma análise de longo prazo das tendências e padrões das dificuldades financeiras. Isso pode ajudar a identificar ciclos econômicos, choques financeiros e crises que afetaram os países latino-americanos ao longo desse período. Além disso, a América Latina enfrentou eventos econômicos significativos, como a crise financeira global de 2008, a crise da dívida soberana na Europa, flutuações nos preços das *commodities*, mudanças políticas e outras influências que impactaram a estabilidade financeira da região durante essas duas décadas. Considerar esses eventos passados pode fornecer *insights* sobre como a região lida com dificuldades financeiras em diferentes cenários.

No que se refere a questão da modelagem, dados financeiros de 2000 a 2020 estão disponíveis e acessíveis para análise em quantidade necessária e

suficiente, o que seria duvidoso com um período menor, por exemplo 10 anos – que desconsideraria a crise financeira global – ou mesmo um período maior, no qual os dados são mais raros, já que tal mercado não possui tantas empresas antes dos anos 2000.

O Brasil é claramente o maior desses mercados, seguido pelo México que, apesar de ter menos empresas listadas, seu valor de mercado supera o Chile. Outros países da América Latina (fora da Tabela 1) não foram incluídos por falta de dados disponíveis ou por conterem poucos dados.

Para realizar análises e previsões consistentes, é fundamental contar com dados confiáveis e que tenham a possibilidade de identificar padrões entre os indivíduos, no caso, das empresas. Se num país possui poucos dados disponíveis ou se os dados disponíveis são inconsistentes ou de baixa qualidade, pode ser difícil obter resultados confiáveis e representativos para esse país específico. Ao construir um modelo de previsão para dificuldades financeiras na América Latina, é importante ter uma amostra representativa dos países da região. Se um país tiver poucos dados disponíveis, sua inclusão na amostra pode não ser representativa do cenário geral da região. Portanto, excluí-lo da amostra pode garantir uma representação mais equilibrada dos países latino-americanos.

O número de observações inclui empresas que já foram deslistadas por algum motivo e tiveram dados agrupados por pelo menos dois anos consecutivos antes da deslistagem.

Com a presença de valores inconsistentes, foi necessária a aplicação de critérios de exclusão de dados. Assim, as observações que apresentaram tais características (por exemplo, falta de alguma variável explicativa ou *outliers*) e menos de dois anos consecutivos de informações disponíveis foram excluídas da amostra. Essas escolhas são detalhadas mais a frente.

3.2 Variáveis Dependente e Independentes

A definição de *Financial Distress* (FD) pode apresentar diferentes contextos, como (1) falência (Altman, 1968; Frydman et al., 1985; Gupta et al., 2018), (2) insolvência, (3) incapacidade de cumprir obrigações financeiras e/ou (4) diminuição do valor de mercado (Manzaneque et al., 2016; Pindado et al., 2008) por dois anos consecutivos de perda de lucro líquido (Geng et al., 2015), e (6) queda no patrimônio líquido (Gupta et al., 2018; Keasey et al., 2015, entre outros). Nesta pesquisa, adota-se a definição de Pindado et al. (2008), em que o FD é verificado quando ocorre pelo menos um dos seguintes eventos: EBITDA é menor que despesas financeiras por dois anos consecutivos; e o va-

lor de mercado da empresa diminui monotonamente em um intervalo de dois anos.

Seguindo o pensamento de Acharya et al. (2007) e Gupta et al. (2018), avalia-se mais variantes dessa medida. A primeira considera as empresas que atenderam, pelo menos, a um desses critérios. Esta definição é a mesma usada por Manzaneque et al. (2016). E a última forma de FD é a interseção deles. A Tabela 2 apresenta essas definições. É importante mencionar que as empresas que foram classificadas como FD também aparecem na amostra classificadas como solventes, mas em anos que essas condições não foram atendidas. Assim, uma empresa pode ter vivenciado mais de um caso de estresse, consecutivos ou não, bem como períodos de solvência antes e depois de estar em um período (ou mais) de dificuldade.

Tabela 2: Definições de FDs usadas para realizar este trabalho e o número de casos identificados no banco de dados.

Variável	Definição	Ocorrências de FD (empresa-ano)
FD1	EBITDA menor que despesas financeiras por 2 anos seguidos	553
FD2	Queda no valor de mercado por 2 anos seguidos	3869
FD3	FD1 ou FD2	4155
FD4	FD1 e FD2	267

As variáveis independentes foram selecionadas levando em consideração dois critérios: relevância em estudos relacionados (Altman, 1968; Altman et al., 1979, 2017; Chen, 2011; Kim & Upneja, 2014; Gupta et al., 2018; López-Gutiérrez, Sanfilippo-Azofra, & Torre-Olmo, 2015; Manzaneque et al., 2016; Rezende et al., 2017; Sayari & Mugan, 2017, entre outros) e disponibilidade no banco de dados. A Tabela 3 mostra todas as variáveis utilizadas nesta pesquisa e uma breve descrição.

Tabela 3: Variáveis Independentes utilizadas como fonte de informação para os modelos previsores e suas respectivas definições.

Variável	Descrição	Fórmula	Referência
Country	País da sede da empresa		López-Gutiérrez et al. (2015)
CPB	Variação no *Price-to-Book* (P/B) entre dois anos consecutivos	$P/B_t - P/B_{t-1}$	Barboza et al. (2017)
CROE	Variação no Retorno sobre o Patrimônio Líquido (ROE) entre dois anos consecutivos	$ROE_t - ROE_{t-1}$	Barboza et al. (2017)
DTE	Dívida Total (TD) sobre o Patrimônio Líquido (PL)	TD / PL	Barboza et al. (2017)
EPS	Lucro por ação	$\frac{\text{Lucro Líquido}}{\text{Quantidade de Ações}}$	Chen (2008)
GSA	Crescimento de Receitas entre dois anos consecutivos	$\frac{\text{Receitas}_t}{\text{Receitas}_{t-1}} - 1$	Barboza et al. (2017)
LIQ	Liquidez Capital de Giro (CG) sobre o Passivo Circulante (PC)	CG / PC	Gupta et al. (2018)
LVC	Composição da dívida: Passivo Circulante (PC) por TD	PC / TD	Rezende et al. (2017)
NICA	Lucro Líquido por Ativo Circulante	LL / AC	Chen (2011)
NIE	Lucro Líquido para o Patrimônio Líquido	LL / PL	Rezende et al. (2017)
OPM	Margem Operacional	$\frac{\text{Lucro Operacional}}{\text{Receitas}}$	Barboza et al. (2017)
Sector	Setor Econômico (Thomson Reuteurs)		Rezende et al. (2017)
X1A	Medida de liquidez de Altman	CG / AT	Altman (1968)
X2A	Medida de lucratividade de Altman	LR / AT	Altman (1968)
X3A	Medida de eficiência operacional de Altman	EBIT / AT	Altman (1968)
X4A	Medida relacionada ao mercado de Altman	VM / VP	Altman (1968)
X5A	Medida de giro de ativos de Altman	Receitas / AT	Altman (1968)

3.3 Procedimento e Modelagem de Previsão

A amostra foi então dividida em dois subconjuntos. A primeira parte, denominada amostra de treinamento, foi criada a partir da seleção de 70% dos dados mais antigos disponíveis, sendo o restante utilizado para verificar o poder preditivo dos modelos, também conhecido como amostra de teste.

Devido à ocorrência de dados desbalanceados, reduziu-se as observações de empresas solventes, por seleção aleatória. Essa etapa forneceu uma amostra balanceada, apropriada para a execução de modelos, principalmente no caso logit.

Foram utilizadas duas técnicas de modelagem: regressão logística (LR), como benchmark, e Random Forest (RF), como proposta baseada em resultados superiores em relação ao LR em estudos semelhantes, como Barboza et al. (2017), e Halteh et al. (2018).

A escolha de uma segunda técnica baseia-se na ineficiência do LR com dados financeiros, conforme explicado por Khemakhem & Boujelbene (2018). Outro fato que contribui para a escolha do RF é a possibilidade de examinar as variáveis explicativas em termos de relevância para a classificação da empresa como FD, o que também foi realizado por Geng et al. (2015).

O LR pretende prever com probabilidade a ocorrência de um evento, considerando a combinação linear das variáveis independentes. Essa combinação é então usada como base para a função logística, que fornece a probabilidade de inadimplência como resultado. Em termos matemáticos,

$$P_{i,t}(Y_{i,t+1} = 1) = \frac{1}{1 + e^{-\alpha - \beta X_{i,t}}}, \tag{2}$$

onde $P_{i,t}$ é a probabilidade de que a empresa i terá dificuldades financeiras um certo tempo t a frente. $Y_{i,t}$ é uma variável binária que assume o valor de 1 (0) se a empresa esteve em FD (não FD) no período $t + 1$. $X_{i,t}$ é o vetor de n variáveis explicativas conhecidas ao final do período t. α e β representam os parâmetros intercepto e de inclinação que caracterizam a função logística, respectivamente, e são estimados via máxima verossimilhança. Um valor maior de $\alpha + \beta X_{i,t}$ indica maior probabilidade de ocorrer FD.

A vantagem de aplicar o LR é a possibilidade de identificar o efeito de cada variável independente na predição (Avino et al., 2019), além de ser fácil de interpretar (Klepac & Hampel, 2017) e requer menos condições, como a premissa de normalidade dos dados (Inekwe et al., 2018).

Em particular, o LR oferece o potencial de uma empresa entrar em FD

(Pindado et al., 2008). Por outro lado, aspectos não lineares são mais difíceis de entender (Khemakhem & Boujelbene, 2018; Li et al., 2019).

A RF é uma técnica de ensemble que utiliza um algoritmo de aprendizado supervisionado e também é entendida como uma coleção de árvores de decisão (Halteh et al., 2018) para que as variáveis e observações que constroem cada árvore sejam escolhidas aleatoriamente.

Conforme explicado por Bubna et al. (2020), a técnica de RF é baseada em n árvores em que cada nó é uma variável escolhida aleatoriamente do banco de dados e o desenvolvimento da árvore também requer k observações extraídas aleatoriamente do banco de dados. Assim, a árvore fornece uma classificação única para todas as observações. O valor de saída representa a escolha dada pela maioria das árvores (Koutanaei et al., 2015).

Assim como LR, Halteh et al. (2018) lembram que este método pode ser usado tanto em problemas de regressão quanto de classificação. A RF apresenta vantagens interessantes: (1) como as variáveis mudam entre as árvores, o método tende a se afastar das correlações; (2) como a votação final depende de vários subconjuntos, o problema de *overfitting* pode ser geralmente descartado; (3) outliers não causam complicações porque a estrutura das árvores de decisão é robusta para isso (Barboza et al., 2017); e (4) outro recorte relevante é a importância das variáveis, o que pode possibilitar uma análise complementar sobre os determinantes (Barboza et al., 2017; Geng et al., 2015; Klepac & Hampel, 2017); e (5) em suma, as árvores estão disponíveis para o espectador, oferecendo a oportunidade de retratar toda a floresta (Klepac & Hampel, 2017).

Realizou-se ainda um processo de validação cruzada, que é um elemento central de verificação da estabilidade do modelo, melhoria de desempenho (Frydman et al., 1985) e teste de robustez (Geng et al., 2015). Esta etapa permite que os modelos de previsão ajustem seus parâmetros da melhor forma, resultando em melhores resultados (menores erros, maior precisão – ponderados ou não).

Em relação ao horizonte de tempo para a previsão de FD, avaliou-se três níveis: um, dois e três anos antes de atingir o status de estresse, como feito por Chen (2008); Gupta et al. (2018), e Klepac & Hampel (2017). Outras possibilidades foram avaliadas, mas a limitação do banco de dados latino-americano impediu o uso de intervalos maiores. A Figura 1 elucida a porção de observações para cada um dos 4 tipos de FD por horizonte de tempo.

Embora a distância do FD aumente, o número de casos aumenta ligei-

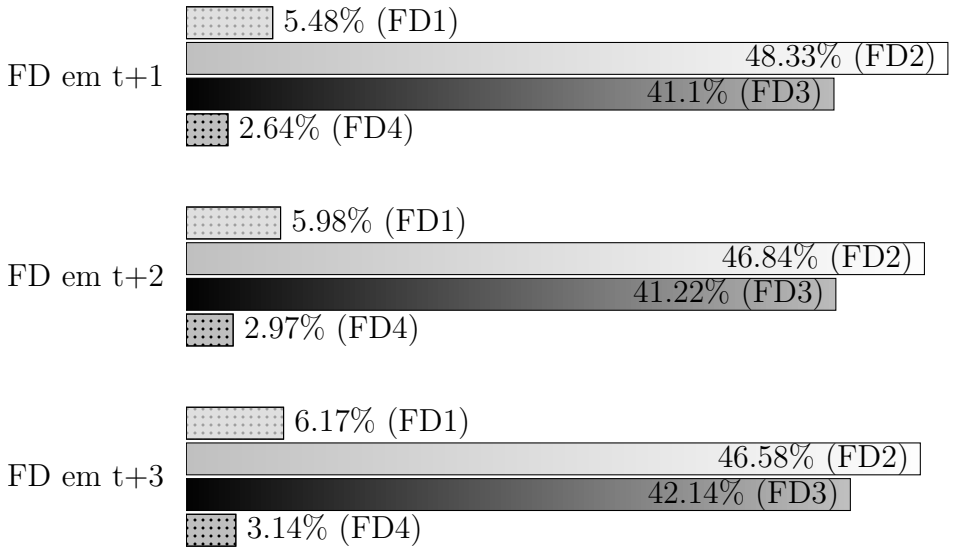


Figura 1: Gráfico de barras mostrando a proporção de casos de FD na América Latina entre 2000 e 2020, considerando quatro definições específicas de FD mostradas na Tabela 2, excluídos os valores faltantes.

ramente para FD1, FD3 e FD4, o que pode ser justificado pela extração de valores ausentes e outliers após a identificação de FD.

Como os casos de FD2 e FD3 exigem a condição com base na visão de mercado, seu percentual expressivamente maior não surpreende. Em contrapartida, a baixa ocorrência de casos FD1 e FD4 pode ser prejudicial para o desenvolvimento de modelos preditivos, uma vez que podem ser considerados casos raros.

3.4 Método de Análise

As medidas avaliadas para comparar o poder preditivo são: AUC, ACC, Erros Tipo I e II, e Brier Score (BS, que é semelhante à medida de erro quadrático médio). A matrix de confusão, dada genericamente pela Tabela 4, é o ponto de partida de algumas delas.

Tabela 4: Representação da Matrix de Confusão. Estrutura básica que resume todos os resultados das previsões em uma única tabela.

	Real	
Previsão	Não FD (0)	FD (1)
Não FD (0)	Verdadeiro Positivo (VP)	Falso Positivo (FP)
FD (1)	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Assim, é possível estimar as métricas ACC, Erros do Tipo I e II, aplicadas em vários estudos para avaliar a qualidade das previsões. Tais indicadores podem ser representados por:

$$\text{ACC} = \frac{VP + VN}{VP + FP + FN + VN} \tag{3}$$

$$\text{Erro do Tipo I} = \frac{FP}{FP + VN} \tag{4}$$

$$\text{Erro do Tipo II} = \frac{FN}{FN + VP} \tag{5}$$

AUC e BS são mais aconselhadas nesse tipo de estudo, pois a amostra de teste possui dados desbalanceados (proporção amplamente maior de empresas solventes do que casos insolventes) como feito por Avino et al. (2019); Khemakhem & Boujelbene (2018); Klepac & Hampel (2017); Mousavi & Ouenniche (2018), entre outros. Além disso, BS capta uma dimensão do erro em relação a probabilidade e não somente se a previsão está correta ou não. Isso pode ser percebido na representação matemática, a seguir:

$$\text{BS} = \frac{1}{N}\sum_{t=1}^{n}(p_t - o_t)^2, \tag{6}$$

onde p_t é a probabilidade de FD dada pelo modelo e o_t é o valor observado, ou seja, $o_t = 0$ se a empresa foi solvente ou $o_t = 1$ se a empresa teve dificuldades financeiras no futuro.

A utilização dos dados desbalanceados é uma escolha que considera uma visão mais realista do fenômeno, já que empresas saudáveis são mais comuns, como pode ser visto em Figura 1 e Tabela 5 da estatística descritiva mostrada a seguir nos resultados.

Os resultados foram então agrupados, analisados e comparados com outros estudos, considerando as quatro formas da variável dependente, os três períodos de previsão e as cinco medidas de desempenho.

Uma das saídas interessantes dos modelos é a análise das variáveis em termos de qualidade para maior assertividade nas previsões. Assim, essa análise também é realizada e discutida nesta tese.

4 Resultados e Análises

O contexto permite que diversos aspectos sejam discutidos, mas realiza-se nesta seção uma visão geral em torno do poder de previsão de cada técnica em termos de previsões de curto, médio e longo prazo, bem como uma anatomia crucial sobre as melhores variáveis preditoras com as quais é feita uma análise do impacto, se houver, da pandemia em tais descobertas.

Primeiramente, mostra-se as estatísticas descritivas das variáveis explicativas (Tabela 5). Em suma, a comparação nesses dois aspectos (mediana e variabilidade) indica que a discriminação entre os dois grupos não é aparentemente fácil de ser observada, pois são semelhantes, com poucas exceções.

Por um lado, são apenas dois indicadores em observação, o que torna precipitado afirmar que são dois grupos com certo grau de parentesco. Por outro lado, pode-se observar um padrão interessante: variáveis com medianas negativas (somente empresas em dificuldades) apresentaram desvios-padrão maiores do que empresas consideradas solventes.

Tabela 5: Estatística Descritiva das Observações Empregadas no Estudo. Neste caso, considera-se as empresas em dificuldades como FD1 (ou seja, pelos fundamentos das empresas).

Variável	Empresas com FD		Empresas sem FD	
	Mediana	Desvio-Padrão	Mediana	Desvio-Padrão
CPB	0,0000	3,97	0,0000	2,95
CROE	-0,1158	1,54	0,0013	1,10
DTE	0,7825	1,78	0,5626	1,03
EPS	-0,2243	10,4	0,7277	4,74
GSA	-0,0786	0,79	0,0725	0,59
LIQ	0,0541	1,57	0,4699	2,06
LVC	1,2120	7,51	0,9062	28,3
NICA	-0,2366	1,06	0,1182	0,50
NIE	-0,2086	1,14	0,0876	0,34
OPM	-0,0735	0,87	0,1120	0,33
X1A	0,0112	0,25	0,0949	0,17
X2A	-0,0586	0,37	0,1457	0,22
X3A	-0,0299	0,09	0,0663	0,08
X4A	0,4497	4,89	1,5121	6,98
X5A	0,3732	0,50	0,5956	0,51
n obs.	553		9.546	

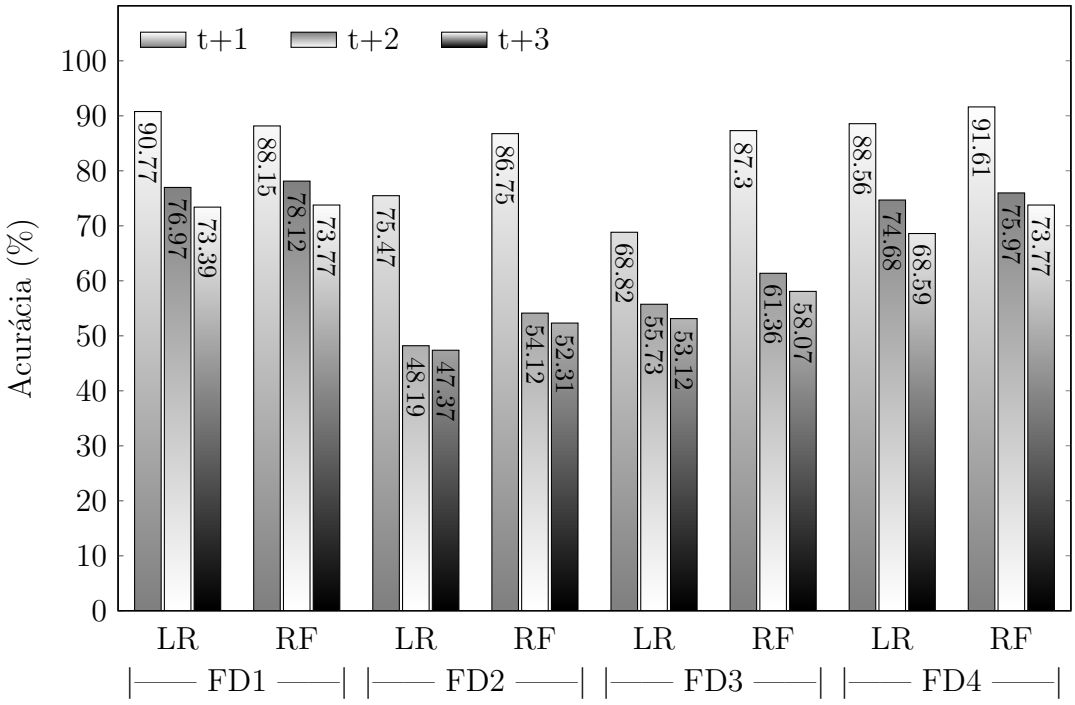
Os modelos foram desenvolvidos após a exclusão de indicadores de correlação superiores a 0,5, o que ocorreu com 4 variáveis. Este procedimento é

indicado por Chen (2008) para evitar multicolinearidade. As tentativas com todas as variáveis não revelaram melhorias e, em seguida, desconsidera-se suas saídas na continuidade do estudo.

4.1 Previsões ao longo do tempo

Os modelos de previsão foram desenvolvidos utilizando 70% das observações para amostra de treinamento e testados com os 30% restantes, ou seja, o poder preditivo foi verificado a partir de novos dados, para que os modelos sejam validados empiricamente. A Figura 2 mostra a capacidade preditiva para os 3 horizontes de tempo em termos de acurácia média, dada pelo total de classificações corretas em relação a todas as observações testadas.

Figura 2: Gráfico de barras comparando a acurácia dos modelos de previsão um ano (primeira barra), dois anos (segunda) e três anos à frente a divulgação dos resultados das empresas com base nos métodos Regressão Logística (LR) e Random Forest (RF) para cada um dos as formas de variáveis dependentes (FD1 a FD4).



Nesse aspecto, pode-se observar que os modelos baseados em RF são superiores aos modelos LR – especialmente para FD3 – com poucas exceções e desempenho semelhante em quase todos. No entanto, FD2 e FD3 têm níveis de precisão razoavelmente próximos para cada distância prevista.

A queda no poder preditivo ao longo dos anos é esperada (é naturalmente mais fácil prever o que acontecerá em um ano do que em dois anos). Apesar dessa tendência nos níveis de precisão, os modelos de RF não perdem tanta

qualidade à medida que o horizonte de tempo aumenta, o que é mais evidente para FD1.

Esses achados contrastam parcialmente com os modelos baseados na Análise Discriminante de Altman (1968), que teve uma diferença maior que 20% de um ano para dois anos à frente. Isso ocorre possivelmente porque na amostra em questão é muito maior com mais variáveis e a convergência do modelo é mais consistente.

Em oposição, os modelos para FD2 e FD3 mostraram esse comportamento entre 1 e 2 anos (permanecendo um desempenho ruim para 3 anos à frente, próximo a 50%), revelando o quão difícil é conhecer o futuro sobre retornos negativos em um cenário de longo prazo .

Além disso, pode-se observar na Figura 2 que a maior discrepância nos níveis de acurácia entre o modelo RF e LR para FD3 está em um ano anterior ao sofrimento.

Tal desempenho – considerado excelente – pode ser enganoso, pois dados desbalanceados podem apresentar diferenças expressivas no número de observações em cada classe, como ocorreu em Barboza et al. (2017).

Dessa forma, um modelo de previsão pode simplesmente expressar alta precisão mesmo classificando todos os casos em um único grupo, o que significa que não há necessidade de um modelo. Assim, a análise dos erros contribui para a identificação da qualidade da previsão. Tabela 6 mostra os erros calculados na amostra de teste.

No ano de previsão, o método RF apresentou acurácia muito maior para FD2, mas o mesmo não ocorre para os demais. É interessante notar que os resultados, apresentados em Tabela 6 não confirmou os contornos de Figura 2, pois RF parece ser muito melhor que LR quando o principal indicador é o erro Tipo II.

Mais especificamente, as taxas de erro do Tipo II de RF são impressionantes quando comparadas a LR e surpreendentemente para o horizonte curto (t+1), onde RF previu todas as empresas em dificuldades para FD1. Tais níveis de erro também são superiores a literatura com proposta semelhante, como Barboza et al. (2017).

Ainda assim, a melhor acurácia do modelo LR apresentou taxas de erro relativamente baixas e desempenho superior ao RF em termos de acurácia, o que torna este modelo atrativo neste horizonte de tempo, o que é consistente com Chen (2011).

Além disso, a qualidade para estimar a probabilidade de ser um caso não

Tabela 6: Erros Tipo I e II dos Modelos Preditivos baseados em LR e RF para cada Horizonte de Previsão. O erro do tipo I mostra a proporção de empresas que solventes de fato, mas o modelo classificou como FD, enquanto o erro do tipo II mostra a porcentagem de empresas FD classificadas como solventes.

FD /	Previsões para t+1		Previsões para t+2		Previsões para t+3	
Erro	LR	RF	LR	RF	LR	RF
FD1						
Tipo I	8.63%	12.52%	22.55%	22.19%	26.41%	26.34%
Tipo II	19.88%	0%	30.95%	16.67%	29.61%	24.58%
FD2						
Tipo I	29.97%	14.56%	50.25%	50.5%	54.03%	48.96%
Tipo II	19.55%	12.06%	53.28%	41.52%	51.3%	46.49%
FD3						
Tipo I	29.69%	15.48%	44.36%	52.09%	45.97%	50.32%
Tipo II	32.97%	9.36%	44.16%	22.6%	47.92%	32.34%
FD4						
Tipo I	11.34%	8.44%	25.21%	24.44%	31.72%	25.66%
Tipo II	14.94%	6.9%	28.89%	11.11%	22.45%	26.53%

angustiado é clara em um horizonte mais longo (t+3), uma vez que as taxas de erro do Tipo I estão sensivelmente próximas das estimativas de erro para t+2, qualquer que seja a definição de angústia .

Em geral, esses resultados estão de acordo com a literatura, como Barboza et al. (2017), e Kim & Upneja (2014), que apresentaram o método de RF como o mais adequado em todos os casos estudados.

De fato, ao se examinar apenas a precisão, o LR supera o RF para classificar uma empresa como financeiramente em dificuldades ou não com base em fundamentos (FD1), mas apenas um ano antes. No entanto, quando o analista está interessado na probabilidade de inadimplência, o RF é inequívoco.

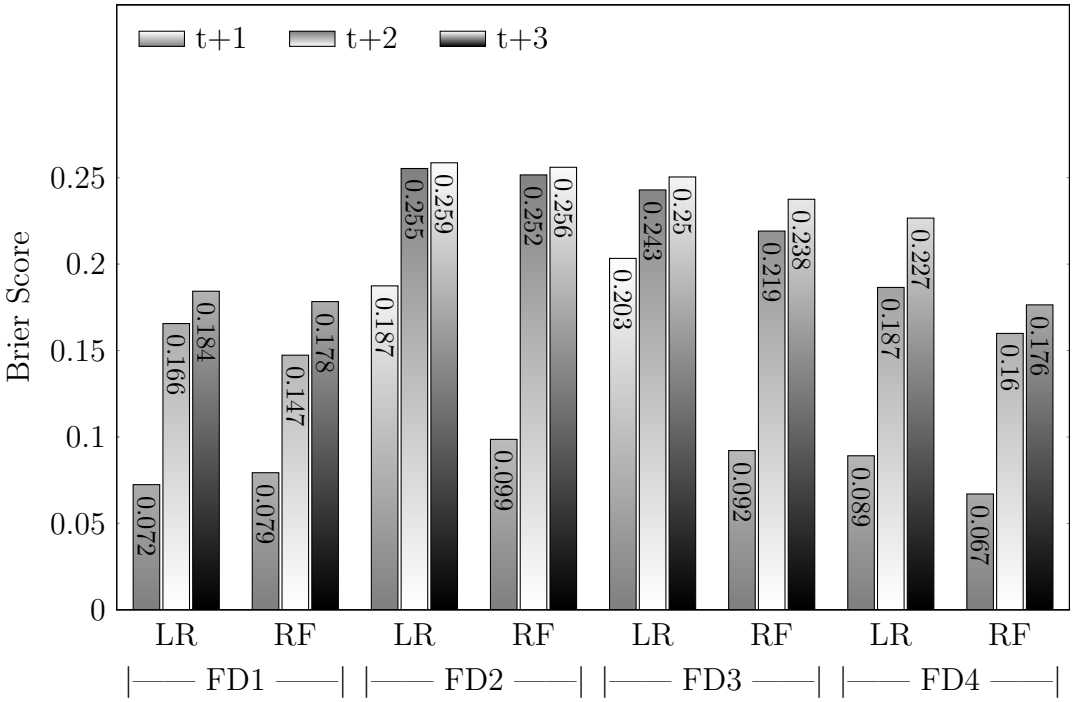
No entanto, deve-se reconhecer que os resultados para FD2 em médio e

longo prazos são fracos. Consequentemente, esses resultados não são discutidos aqui e pode-se concluir que vislumbrar retornos negativos até o momento não é uma tarefa fácil, principalmente quando se utiliza indicadores financeiros como variáveis explicativas.

Duas outras medidas relacionadas à previsão fornecem informações adicionais a esses resultados. O Brier Score (BS) é um indicador que mede o erro entre o valor previsto e o valor observado, que também foi aplicado por Mousavi & Ouenniche (2018). Então, quanto menor, melhor.

A segunda medida é a área na curva ROC (AUC), amplamente utilizada para medir a virtude dos modelos testados em dados não balanceados (Barboza et al., 2017). Figura 3 mostra uma comparação para os valores de BS e Figura 4 apresenta o gráfico comparativo das curvas ROC.

Figura 3: Brier Score para cada modelo, considerando o horizonte de previsão e o tipo de dificuldade (FD1–FD4).

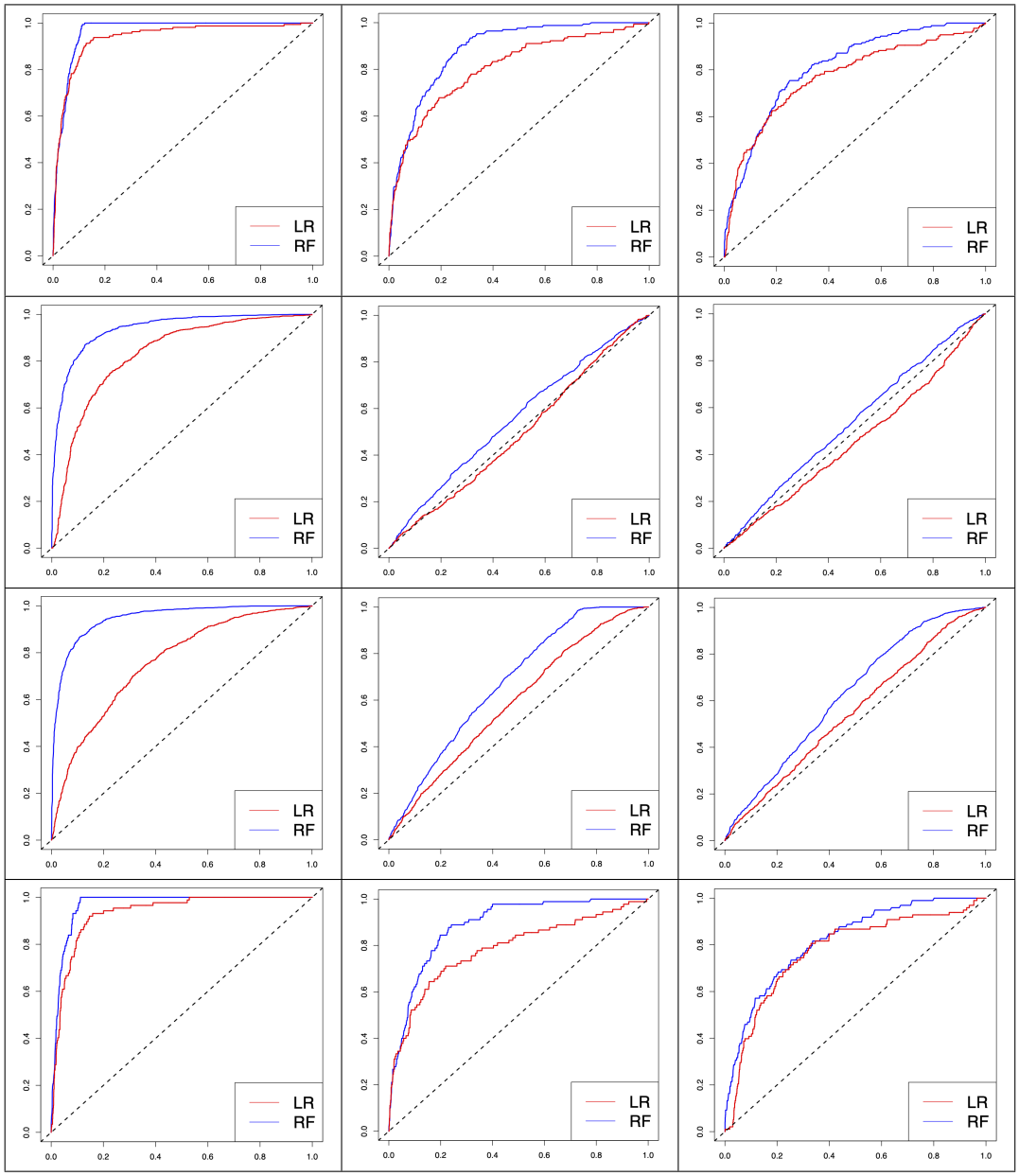


Os níveis de BS mostram que os modelos RF são, em média, mais discriminativos do que LR. O melhor modelo, considerando esta medida, é aquele utilizado para prever FD4 com um ano anterior ao distress.

Outra coisa demonstrada na Figura 3 é a involução em BS sobre as definições de FD. Talvez isso venha como o incremento de restrições de FD1 para FD4. Além disso, a regressão logística não é qualificada para predizer FD3.

As curvas ROC mostram a supremacia dos modelos RF em relação aos

Figura 4: Curvas ROC (Falso Positivo x Verdadeiro Positivo). As colunas representam o horizonte de previsão para FD (ou seja, t+1, t+2 e t+3, respectivamente). Cada linha é fornece o tipo de FD (FD1 até FD4).



modelos LR, o que é claramente representado na Figura 4, onde a maioria dos gráficos, as curvas dos modelos RF são curvas LR bastante superiores.

Ao avaliar apenas esta medida, a predição em t+1 pela técnica de RF é bastante satisfatória (cerca de 0,9) e, com isso, confirma o melhor desempenho do RF em relação ao LR.

Como resultado, não se tem dúvidas de que o RF superou o LR para prever dificuldades financeiras de empresas públicas latino-americanas em 1, 2 ou 3 anos à frente.

4.2 Classificação de Importância das Variáveis Independentes

Para complementar a análise feita nesta pesquisa, é feito um estudo sobre a importância de cada variável. Deve-se notar que os modelos RF e LR são diferentes a este respeito. Figura 5 revela a variável explicativa mais importante para prever FD, técnica de modelagem, horizonte de tempo e entre modelos LR e RF.

Os modelos de RF mostraram que as medidas relacionadas aos lucros são variáveis consistentemente importantes. Em particular, a influência do EPS também foi verificada e sublinhada por Geng et al. (2015).

As medidas de crescimento também tiveram uma adequação considerável para esses modelos. De acordo com Gepp & Kumar (2015), e Laitinen & Suvas (2016), razões considerando variabilidade forneceriam informações relevantes para as previsões.

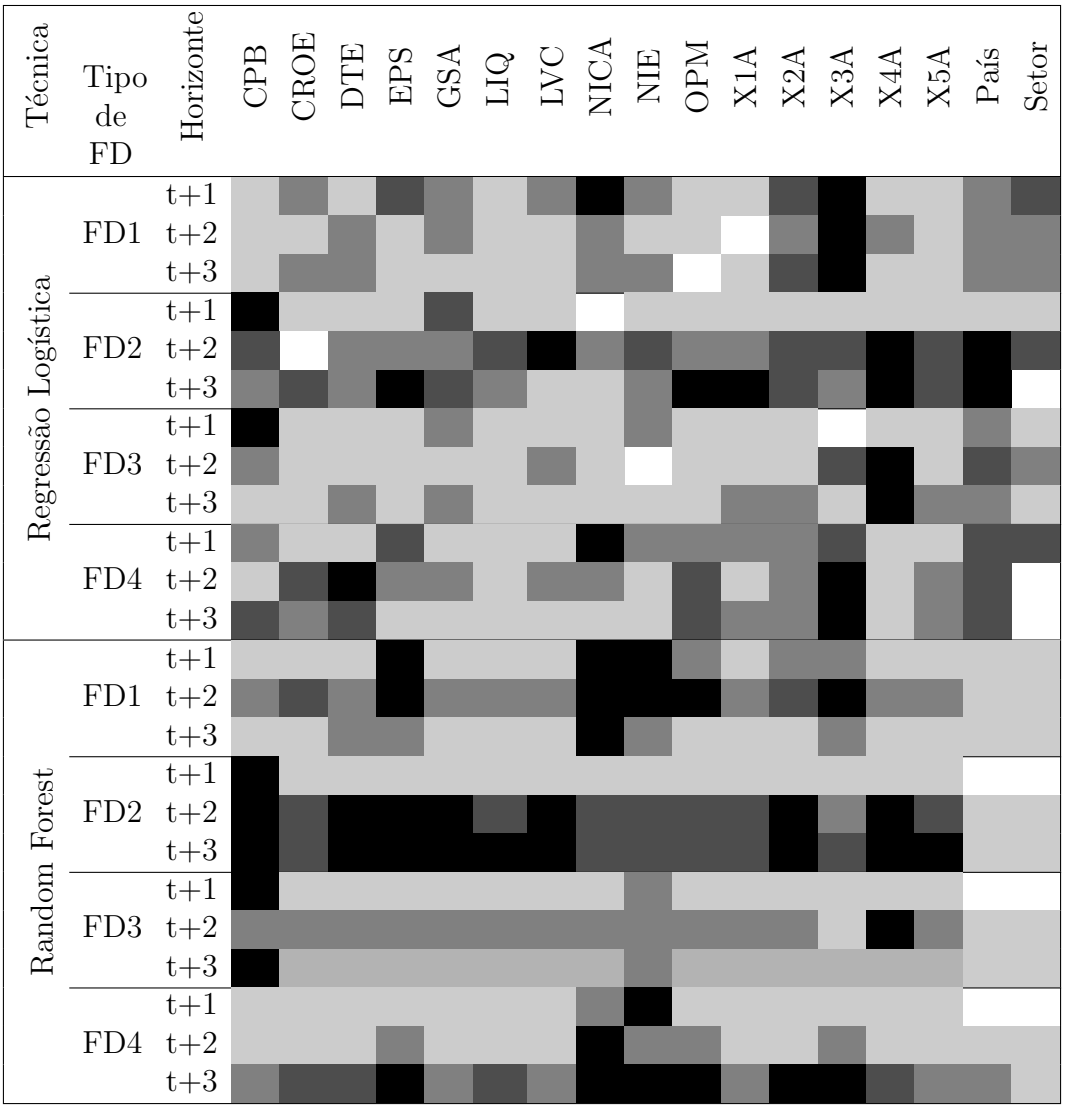
Além disso, Garmaise & Natividad (2021) destacou a ligação entre crescimentos mais altos e angústia. A CEC revela uma forte importância da RF em FD1 e FD2 em todos os horizontes. A possível razão para isso é a capacidade dos investidores em ajustar o preço das ações para empresas em dificuldades.

Quase todos os modelos indicam que as medidas do Altman (1968) são relevantes. Além disso, os indicadores de crescimento e mudança também são de importância relativa, mostrando que tais achados estão alinhados ao Barboza et al. (2017).

Nos modelos de RF, as medidas relacionadas aos lucros podem ser consideradas uma contribuição proeminente para a previsão de FD, notadamente EPS e NICA.

As variáveis de Altman também revelaram sua importância relativa considerando a qualidade dos índices financeiros das firmas. Independentemente da

Figura 5: Mapa de calor que reflete a Importância de cada variável explicativa para ambas as técnicas de modelagem e todos os horizontes de previsão. Preto significa um melhor preditor (ou determinante) e branco pode ser considerado irrelevante.



técnica empregada, especialmente X3A e X4A parecem ser consistentemente muito importantes.

As variáveis de interesse Setor e País tiveram pouca ou nenhuma influência nos modelos de RF, embora sua exclusão cause perda de qualidade, ou seja, uma ou mais medidas de desempenho relataram queda no poder de predição do modelo.

No entanto, quase todos os modelos de LR consideraram País e Setor como variáveis fundamentais, especialmente para FD1 e FD2.

Esses resultados sugerem que os comportamentos não lineares são majoritariamente apresentados em índices financeiros, conforme observado por Li et al. (2019) uma vez que os modelos RF capturam essa propriedade melhor do que LR.

Escobari et al. (2017) encontrou forte relação entre seis países da América Latina em períodos entre eventos de bolhas e Keasey et al. (2015) afirmou que o FD pode ser influenciado por condições específicas de cada país na Europa.

Em relação ao Setor, Altman et al. (2017) lembra do potencial do X5A para ser afetado pela indústria da empresa, e então sugeriu um modelo sem ele.

Não foram encontrados mais estudos que discutiam as diferenças de setores e sua relação com o FD. Como LR e RF são diferentes em suas concepções, não há consenso na literatura sobre essas duas características, em especial na América Latina, pode-se constatar que País e Setor são relevantes no modelo Logit, mas irrelevantes para RF.

4.3 Análise de persistência no tempo e o efeito da pandemia

O início da pandemia do COVID-19 trouxe à tona notícias sem precedentes para a sociedade moderna em todos os lugares do mundo. Nesse sentido, as previsões para 2020 intuitivamente são vistas com desconfiança por investidores, profissionais e gestores.

Pensando nisso, examina-se aqui o poder de previsão da RF neste ano e comparando com as previsões anteriores, para perceber se esse enorme efeito pode causar incrementos nas taxas de erro. As Tabelas 7 e 8 expressam as apurações em dois cenários de previsões (antes e depois do início da pandemia) com um teste estatístico para confirmar a diferença entre elas.

Tabela 7: Desempenho de RF com relação a acurácia antes e depois do início da pandemia. Nível de Significância: (*)10%, (**)5%, e (***)1%

		Acurácia	
Horizonte	FD	-2019	2020
t+1	FD1	2135 (88%)	521 (87%)
	FD2	1691 (86%)	456 (89%)*
	FD3	2091 (87%)	541 (90%)**
	FD4	2222 (92%)	540 (90%)
t+2	FD1	1820 (77%)	483 (82%)***
	FD2	1070 (54%)	263 (53%)
	FD3	1473 (62%)	337 (57%)**
	FD4	1777 (75%)	464 (79%)
t+3	FD1	1697 (73%)	432 (75%)
	FD2	1073 (54%)	219 (44%)***
	FD3	1369 (59%)	308 (53%)**
	FD4	1707 (74%)	439 (76%)

Tabela 8: Desempenho de RF com relação aos erros antes e depois do início da pandemia. Nível de Significância: (*)10%, (**)5%, e (***)1%

		Erro do Tipo I		Erro do Tipo II	
Horizonte	FD	-2019	2020	-2019	2020
t+1	FD1	280 (12%)	77 (14%)	0 (0%)	0 (0%)
	FD2	146 (16%)	26 (10%)**	127 (12%)	29 (12%)
	FD3	215 (16%)	40 (12%)**	110 (10%)	18 (7%)
	FD4	190 (8%)	57 (10%)	4 (6%)	2 (13%)
t+2	FD1	516 (23%)	101 (18%)***	22 (15%)	6 (23%)
	FD2	469 (50%)	135 (51%)	427 (41%)	99 (43%)
	FD3	637 (51%)	199 (57%)**	249 (23%)	55 (23%)
	FD4	574 (25%)	125 (22%)*	8 (10%)	2 (17%)
t+3	FD1	574 (27%)	139 (25%)	39 (26%)	5 (17%)
	FD2	424 (45%)	165 (62%)***	479 (46%)	110 (49%)
	FD3	585 (49%)	190 (57%)**	357 (32%)	79 (33%)
	FD4	582 (26%)	134 (24%)	22 (26%)	4 (29%)

É possível ver que as previsões não tiveram impacto estatisticamente significativo pelo COVID19 em 32 das 45 medidas, especialmente para erros do Tipo II. Assim, conclui-se que apesar de uma variedade de eventos ocorridos em 2020, RF apresentou o mesmo desempenho para prever FD, atingindo pelo menos 3 de 4 previsões corretas no início da pandemia e mantendo uma taxa de reconhecimento de FD em torno de 88% em um ano antes.

Alguns desses resultados merecem ser enfatizados. Primeiramente, RF não errou previsão alguma de empresas em FD1 um ano antes do evento, mesmo no período de COVID-19. Tal performance simplesmente torna o modelo perfeito para evitar perdas por inadimplência e, portanto, considera-se uma ferramenta diferenciada para análise de crédito.

Segundo, prever isoladamente a queda no valor de mercado (FD2) mostra-se um desafio, uma vez que previsões mais longas são praticamente ineficazes. Entretanto, A previsão com horizonte de 1 ano apresenta níveis satisfatórios, sendo até estatisticamente significativa a melhoria nas estimativas no início da pandemia (16% versus 10%).

Terceiro, apesar de poucos dados para FD4, as previsões mostraram-se satisfatórias até mesmo em t+3. Isso sugere que o comportamento ruim das empresas e a percepção do mercado negativa juntos possui uma informação relevante para os participantes desse mercado regional.

5 Considerações Finais

A previsão de dificuldades financeiras é um assunto bastante discutido na literatura e sempre apresenta novas perspectivas e contribuições. Isso é justificado por muitas partes interessadas, conforme destacado por Altman et al. (2017).

Este estudo tratou da previsão de Financial Distress (FD) na América Latina, região pouco estudada no passado. Os modelos investigados envolvem 6 países (Argentina, Brasil, Chile, Colômbia, México e Peru). O FD foi avaliado de 4 formas diferentes, e com a utilização de mais de 20 indicadores financeiros como variáveis independentes. Em paralelo, avaliou-se o desempenho de dois métodos que apresentaram bons resultados neste problema: random forest e regressão logística. Os modelos foram avaliados em 3 horizontes temporais, de 1 a 3 anos anteriores ao FD. Os resultados mostraram desempenho superior para a random forest na maioria das situações.

Em relação aos melhores preditores, indicadores relacionados ao Lucro Líquido, Lucro por Ação, medidas de crescimento e variação nos índices *Price-to-Book* desempenharam papéis importantes nas previsões do método RF. Já nos modelos baseados em RF, destacaram-se os indicadores do Z-Score de Altman (1968) bem como as medidas de crescimento. Pode-se perceber que medidas de lucratividade e crescimento são itens que devem ser considerados em modelos de previsão de dificuldades financeiras, principalmente para empresas latino-americanas. Também observa-se que os modelos de RF tiveram desempenho semelhante antes e no início da pandemia, o que confirma a implementação geral desta análise, independente do período de observação.

Pode-se então concluir que os modelos de previsão de FD na América Latina alcançam resultados significativos em intervalos suficientes para desenvolver estratégias de alerta antecipado dentro e fora da empresa para evitar prejuízos para os agentes envolvidos. As descobertas aqui apresentadas têm implicações atraentes tanto para gestores quanto para investidores e, logicamente, para o desenvolvimento de ferramentas de gestão de risco.

Apesar de demonstrar resultados encorajadores, esta pesquisa também apresenta limitações e recomendações para inspeções posteriores. Sugere-se primeiramente que estudos futuros incluam variáveis macroeconômicas, medidas de governança corporativa e, diante de tantas medidas, associe critérios de seleção de variáveis para tornar o modelo mais simples e, dessa forma, ser aplicado por mais praticantes. Além disso, recomenda-se o desenvolvimento de modelos específicos para pequenas e médias empresas, públicas e privadas.

Referências

Acharya, V. V., Bharath, S. T., & Srinivasan, A. (2007). Does industry-wide distress affect defaulted firms? evidence from creditor recoveries. *Journal of Financial Economics*, *85*(3), 787–821.

Alifiah, M. N. (2014). Prediction of financial distress companies in the trading and services sector in malaysia using macroeconomic variables. *Procedia-Social and Behavioral Sciences*, *129*, 90–98.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*(4), 589–609.

Altman, E. I. (2018). Applications of distress prediction models: What have we learned after 50 years from the z-score models? *International Journal of Financial Studies*, *6*(3), 70.

Altman, E. I., Baidya, T. K., & Dias, L. M. R. (1979). Assessing potential financial problems for firms in brazil. *Journal of International Business Studies*, *10*(2), 9–24.

Altman, E. I., Eom, Y. H., & Kim, D. W. (1995). Failure prediction: evidence from korea. *Journal of International Financial Management & Accounting*, *6*(3), 230–249.

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial distress prediction in an international context: A review and empirical analysis of altman's z-score model. *Journal of International Financial Management & Accounting*, *28*(2), 131–171.

Avino, D. E., Conlon, T., & Cotter, J. (2019). Credit default swaps as indicators of bank financial distress. *Journal of International Money and Finance*, *94*, 132–139.

Bali, T. G., Del Viva, L., Lambertides, N., & Trigeorgis, L. (2020). Growth options and related stock market anomalies: Profitability, distress, lotteryness, and volatility. *Journal of Financial and Quantitative Analysis*, *55*(7), 2150–2180.

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems With Applications*, *83*, 405–417.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 71–111.

Bubna, A., Das, S. R., & Prabhala, N. (2020). Venture capital communities. *Journal of Financial and Quantitative Analysis*, *55*(2), 621–651.

Cardona, L., Gutiérrez, M., & Agudelo, D. A. (2017). Volatility transmission between us and latin american stock markets: Testing the decoupling hypothesis. *Research in International Business and Finance*, *39*, 115–127.

Carvalho, D., Ferreira, M. A., & Matos, P. (2015). Lending relationships and the effect of bank distress: evidence from the 2007–2009 financial crisis. *Journal of Financial and Quantitative Analysis*, *50*(6), 1165–1197.

Chen, H. H. (2008). The timescale effects of corporate governance measure on predicting financial distress. *Review of Pacific Basin Financial Markets and Policies*, *11*(01), 35–46.

Chen, M.-Y. (2011). Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems With Applications*, *38*(9), 11261–11272.

Escobari, D., Garcia, S., & Mellado, C. (2017). Identifying bubbles in latin american equity markets: Phillips-perron-based tests and linkages. *Emerging Markets Review*, *33*, 90–101.

FIA. (2021). *Futures industry association – etd tracker.* Disponível em: `https://www.fia.org/fia/etd-tracker`. (Acessado em: 2021-04-13)

Frydman, H., Altman, E. I., & Kao, D.-L. (1985). Introducing recursive partitioning for financial classification: the case of financial distress. *The Journal of Finance*, *40*(1), 269–291.

Garmaise, M. J., & Natividad, G. (2021). Financial flexibility: At what cost? *Journal of Financial and Quantitative Analysis*, *56*(1), 249–282.

Geng, R., Bose, I., & Chen, X. (2015). Prediction of financial distress: An empirical study of listed chinese companies using data mining. *European Journal of Operational Research*, *241*(1), 236–247.

Gepp, A., & Kumar, K. (2015). Predicting financial distress: A comparison of survival analysis and decision tree techniques. *Procedia Computer Science*, *54*, 396–404.

Gupta, J., Gregoriou, A., & Ebrahimi, T. (2018). Empirical comparison of hazard models in predicting smes failure. *Quantitative Finance*, *18*(3), 437–466.

Halteh, K., Kumar, K., & Gepp, A. (2018). Financial distress prediction of islamic banks using tree-based stochastic techniques. *Managerial Finance*.

Inekwe, J. N., Jin, Y., & Valenzuela, M. R. (2018). The effects of financial distress: Evidence from us gdp growth. *Economic Modelling*, *72*, 8–21.

Keasey, K., Pindado, J., & Rodrigues, L. (2015). The determinants of the costs of financial distress in smes. *International Small Business Journal*, *33*(8), 862–881.

Khemakhem, S., & Boujelbene, Y. (2018). Predicting credit risk on the basis of financial and non-financial variables and data mining. *Review of Accounting and Finance*.

Kim, S. Y., & Upneja, A. (2014). Predicting restaurant financial distress using decision tree and adaboosted decision tree models. *Economic Modelling*, *36*, 354–362.

Klepac, V., & Hampel, D. (2017). Predicting financial distress of agriculture companies in eu. *Agricultural Economics*, *63*(8), 347–355.

Koutanaei, F. N., Sajedi, H., & Khanbabaei, M. (2015). A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. *Journal of Retailing and Consumer Services*, *27*, 11–23.

Laitinen, E. K., & Suvas, A. (2016). Financial distress prediction in an international context: Moderating effects of hofstede's original cultural dimensions. *Journal of Behavioral and Experimental Finance*, *9*, 98–118.

Lee, T.-S., & Yeh, Y.-H. (2004). Corporate governance and financial distress: evidence from taiwan. *Corporate Governance: An International Review*, *12*(3), 378–388.

Li, Y., Lu, R., & Srinivasan, A. (2019). Relationship bank behavior during borrower distress. *Journal of Financial and Quantitative Analysis*, *54*(3), 1231–1262.

Liang, D., Lu, C.-C., Tsai, C.-F., & Shih, G.-A. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research*, *252*(2), 561–572.

López-Gutiérrez, C., Sanfilippo-Azofra, S., & Torre-Olmo, B. (2015). Investment decisions of companies in financial distress. *BRQ Business Research Quarterly*, *18*(3), 174–187.

Manzaneque, M., Priego, A. M., & Merino, E. (2016). Corporate governance effect on financial distress likelihood: Evidence from spain. *Revista de Contabilidad*, *19*(1), 111–121.

Masulis, R. W., & Simsir, S. A. (2018). Deal initiation in mergers and acquisitions. *Journal of Financial and Quantitative Analysis*, *53*(6), 2389–2430.

Mousavi, M. M., & Ouenniche, J. (2018). Multi-criteria ranking of corporate distress prediction models: empirical evaluation and methodological contributions. *Annals of Operations Research*, *271*(2), 853–886.

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

Oz, I. O., & Yelkenci, T. (2017). A theoretical approach to financial distress prediction modeling. *Managerial Finance*.

Pindado, J., Rodrigues, L., & de la Torre, C. (2008). Estimating financial distress likelihood. *Journal of Business Research*, *61*(9), 995–1003.

Rezende, F. F., Montezano, R. M. d. S., Oliveira, F. N. d., & Lameira, V. d. J. (2017). Predicting financial distress in publicly-traded companies. *Revista Contabilidade & Finanças*, *28*(75), 390–406.

Romero-Barrutieta, A. L., Bulíř, A., & Rodríguez-Delgado, J. D. (2015). The dynamic implications of debt relief for low-income countries. *Review of Development Finance*, *5*(1), 1–12.

Sayari, N., & Mugan, C. S. (2017). Industry specific financial distress modeling. *BRQ Business Research Quarterly*, *20*(1), 45-62. doi: https://doi.org/10.1016/j.brq.2016.03.003

The World Bank. (2021). *World bank financial database.* Retrieved from `http://data.worldbank.org/indicator/cm.mkt.lcap.cd`. (Accessed: 2021-05-20)

Wanke, P., Barros, C. P., & Faria, J. R. (2015). Financial distress drivers in brazilian banks: A dynamic slacks approach. *European Journal of Operational Research*, *240*(1), 258–268.