



Giovanna Niskier Saadia

**Uso de Técnicas de Machine Learning na Previsão do
Risco de Inadimplência de Alunos em uma Instituição de
Ensino Superior Privada**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Administração de Empresas do Departamento de Administração da PUC-Rio.

Orientador: Prof. Jorge Brantes Ferreira

Rio de Janeiro

Abril de 2020



Giovanna Niskier Saadia

**Uso de Técnicas de Machine Learning na Previsão do
Risco de Inadimplência de Alunos em uma Instituição de
Ensino Superior Privada**

Dissertação apresentada como requisito parcial para
obtenção do grau de Mestre pelo Programa de Pós-
Graduação em Administração de Empresas da PUC-Rio.
Aprovada pela Comissão Examinadora abaixo.

Prof. Jorge Brantes Ferreira

Orientador

Departamento de Administração - PUC-Rio

Prof. Marcus Wilcox Hemais

Departamento de Administração - PUC-Rio

Prof. Angilberto Sabino de Freitas

Unigranrio

Rio de Janeiro, 22 de Abril de 2020

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

Giovanna Niskier Saadia

Graduou-se em Administração de Empresas na Pontifícia Universidade Católica do Rio de Janeiro, em 2016.

Ficha Catalográfica

Saadia, Giovanna Niskier

Uso de técnicas de machine learning na previsão do risco de inadimplência de alunos em uma instituição de ensino superior privada / Giovanna Niskier Saadia ; orientador: Jorge Brantes Ferreira. – 2020.

81 f. : il. color. ; 30 cm

Dissertação (mestrado)—Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Administração, 2020

Inclui bibliografia

1. Administração - Teses. 2. Machine learning. 3. Previsão. 4. Inadimplência. 5. Educação superior. 6. Credit scoring. I. Ferreira, Jorge Brantes. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Administração. III. Título.

CDD: 658

Agradecimentos

Aos meus queridos pais, Celso e Andréa, pelo apoio incondicional e por sempre estimularem que eu e a minha irmã, Gabriela, buscássemos o caminho da educação. Essa conquista é nossa!

Ao meu marido, Felipe, pela parceria, compreensão e amor. Sem você ao meu lado, o caminho teria sido muito mais difícil.

Ao meu orientador, Jorge Brantes, pelo apoio e troca durante todo o processo acadêmico.

Resumo

Saadia, Giovanna Niskier; Ferreira, Jorge Brantes. **Uso de Técnicas de Machine Learning na Previsão do Risco de Inadimplência de Alunos em uma Instituição de Ensino Superior Privada**. Rio de Janeiro, 2020. 81p. Dissertação de Mestrado - Departamento de Administração, Pontifícia Universidade Católica do Rio de Janeiro.

Tão expressiva quanto a curva de crescimento do número de matrículas nas instituições de ensino superior (IES) privadas nos últimos anos é a respectiva curva da inadimplência, cujo aumento pode ser explicado, principalmente, pelo aprofundamento da crise econômica no país e pela redução do número de vagas ofertadas pelo FIES. A inadimplência apresenta-se como um desafio à gestão financeira das instituições de ensino, uma vez que impacta os seus custos operacionais e acaba sendo repassada aos alunos sob forma de aumento de mensalidade. Além disso, a evasão estudantil é também uma das principais consequências da inadimplência, à medida que alunos com dificuldades financeiras acabam por abandonar seus cursos, representando para as instituições de ensino não só uma perda econômica, como também acadêmica e social. As IES, em sua maioria, não utilizam qualquer tipo de técnica de *credit scoring* para prever o risco de seus alunos se tornarem inadimplentes. Nesse sentido, este trabalho apresenta uma metodologia quantitativa para previsão de risco de inadimplência de alunos ativos. Baseado em dados históricos de alunos que estavam inadimplentes ou adimplentes, modelos gerados por algoritmos de *machine learning* foram estimados e comparados. Por fim, os resultados obtidos evidenciaram a relação entre a inadimplência e a variação do valor pago ao longo dos semestres analisados, quantidade média de disciplinas cursadas, natureza empregatícia ao aluno e existência de débitos em semestres anteriores. Com a aplicação dos modelos propostos, as IES seriam capazes de identificar alunos com maior risco de inadimplência e planejar ações preventivas específicas para este grupo.

Palavras-chave

Machine learning; previsão; inadimplência; educação superior; *credit scoring*.

Abstract

Saadia, Giovanna Niskier; Ferreira, Jorge Brantes (Advisor). **The Application of Machine Learning Framework to Identify Students at Risk of Default in a Higher Education Institution**. Rio de Janeiro, 2020. 81p. Dissertação de Mestrado - Departamento de Administração, Pontifícia Universidade Católica do Rio de Janeiro.

As impressive as the growth rate in the number of enrollments in private higher education institutions in recent years is the increase in the related default rate, driven by the deepening economic crisis in Brazil and by the reduction of the number of vacancies offered by the FIES. Default presents itself as a challenge to the financial management of educational institutions, since it impacts their operational costs and ends up being passed on to students in the form of an increase in tuition. In addition, student dropout is also one of the main consequences of default, since students with economic difficulties end up abandoning their courses. Most higher education institutions do not use any type of credit scoring analysis to predict the risk of their students becoming defaulters, failing to understand which factors cause it, and, therefore, refraining from planning preventive actions. Therefore, this study presents a quantitative methodology to predict the default risk of active students. Models generated by machine learning algorithms were analyzed based on a historical database of students who were in or not in default. The results showed a relationship between default and economic, academic and social characteristics of students. Thus, by employing models such as the ones proposed, higher education institutions should be able to identify those students who are at higher risk of defaulting and take specific preventive actions to prevent such an outcome.

Keywords

Machine learning; prediction; default; higher education; credit scoring.

Sumário

1	Introdução	11
1.1.	Contextualização da questão de pesquisa	11
1.2.	Objetivo do estudo	14
1.3.	Relevância do estudo	14
1.4.	Delimitação do estudo	16
1.5.	Estrutura do trabalho	17
2	Referencial Teórico	18
2.1.	Customer Relationship Management	18
2.2.	Gerenciamento de crédito ao consumidor	20
2.3.	Avaliações do risco de crédito	21
2.3.1.	Avaliações subjetivas de crédito	22
2.3.2.	Avaliações objetivas de crédito – <i>credit scoring</i>	24
2.4.	Aplicações de <i>credit scoring</i> em diferentes setores de negócio	25
2.4.1.	Aplicação de <i>credit scoring</i> no setor de educação básica	26
2.4.2.	Aplicação de <i>credit scoring</i> no setor de crédito bancário de pessoas físicas	27
2.4.3.	Aplicação de <i>credit scoring</i> no setor de cartões de crédito	28
2.5.	Inadimplência no setor de educação superior privada	29
2.5.1.	Legislação educacional vigente	31
2.5.2.	Ações contra a inadimplência – políticas de cobrança	32
2.5.3.	Fatores que impactam a inadimplência	34
2.6.	Retenção no setor de educação superior	36
2.6.1.	Fatores-chaves relacionados à retenção de estudantes	38
2.6.2.	Lealdade no setor de educação superior	39
2.7.	Mineração de dados e <i>Machine Learning</i>	40
3	Metodologia	42
3.1	A Instituição de Ensino Superior avaliada	42
3.2	Tipo de Pesquisa	43

3.3	Definição da variável-alvo	44
3.4	Dados coletados	45
3.5	Preparação dos dados	46
3.5.1.	Tratamento e limpeza	46
3.5.2.	Adição de variáveis derivadas	49
3.6	Análise dos dados	51
4	Análise dos resultados	52
4.1	Modelos gerados	52
4.1.1	Acurácia dos modelos	52
4.1.2	Análise do modelo de regressão logística	56
4.1.3	Análise do modelo de árvore de decisão	60
4.1.4	Análise do modelo de redes neurais	64
4.2.	Implicações do uso dos modelos	65
5	Considerações Finais	68
5.1	Oportunidades para futuras aplicações	71
5.2	Limitações do Estudo	72
6	Referencial Bibliográfico	73

Lista de figuras

Figura 4.1 - Estrutura geral de uma árvore de decisão	61
Figura 4.2 - Primeiros níveis da árvore de decisão gerada	62
Figura 4.3 - Estrutura geral de uma rede neural	65

Lista de tabelas

Tabela 3.1 - Tabela de Dados Coletados	45
Tabela 3.2 - Tabela de Variáveis Binárias	47
Tabela 3.3 - Tabela de Variáveis Derivadas	49
Tabela 4.1 – Nomes dos algoritmos para referência	52
Tabela 4.2 - Exemplo de matriz de confusão	53
Tabela 4.3 - Matrizes de confusão dos modelos calculados	54
Tabela 4.4 - Resumo dos resultados de desempenho dos modelos	54
Tabela 4.5 - Atributos e <i>odds ratios</i> estimados pelo modelo de regressão logística	56
Tabela 4.6 – Resumo dos principais resultados dos modelos	67

1 Introdução

1.1. Contextualização da questão de pesquisa

O ensino superior brasileiro registrou, de 2008 a 2018, um crescimento de 44,6% no número de matrículas em cursos de graduação e sequencial, atingindo 8,45 milhões de matrículas em 2018. Deste total, 75,4% na rede privada e 24,6% na rede pública (INEP, 2018). Foi a Lei de Diretrizes e Bases da Educação Nacional (LDB) nº 9.394/1996, que flexibilizou as formas de acesso ao ensino superior, uma vez que as Instituições de Ensino Superior (IES) puderam definir seus próprios mecanismos para admissão de estudantes. Esse crescimento pode ser explicado também pelo avanço dos programas governamentais de concessão de bolsas de estudo parciais e integrais (PROUNI) e crédito financiado (FIES), entre 2007 e 2010, no segundo mandato do governo Luiz Inácio Lula da Silva, que facilitaram o acesso à educação superior para a população de baixa renda. No entanto, tão expressiva quanto a curva de crescimento do número de matrículas, é a curva da inadimplência nas IES privadas, que em 2018 atingiu a taxa de 9,3% das mensalidades com atrasos superiores a 90 dias, permanecendo acima da inadimplência total das pessoas físicas, e apresentando um aumento de 4,5% em relação ao ano anterior (SEMESP, 2019).

Ainda de acordo com o SEMESP (2019), o aumento da inadimplência pode ser explicado pelo aprofundamento da crise econômica no país e de mais uma significativa redução dos contratos novos do FIES. Além dos fatores externos, o esgotamento dos recursos das próprias IES, que possibilitaram a renegociação das dívidas e a ampliação da oferta de crédito estudantil próprio em 2017, também contribuíram para essa elevação da taxa de inadimplência a partir de 2018.

“Inadimplência é a falta de pagamento; inadimplemento é o termo jurídico utilizado, em regra, para designar uma situação de não cumprimento da cláusula contratual; insolvência é a perda total de capacidade de pagamento” (TEIXEIRA; SILVA, 2001, p.19). Para as IES, uma maior taxa de inadimplência significa um aumento de seu custo operacional que acaba sendo repassado aos alunos sob forma

de aumento de mensalidade, fazendo com que, por vezes, alunos adimplentes tornem-se inadimplentes por não conseguirem arcar com o aumento no valor da mensalidade, consolidando um ciclo vicioso (ANDRADE et al, 2008). Desta forma, é importante para as IES entenderem os fatores que levam ao aluno deixar de pagar uma ou mais mensalidades, para que possam tomar ações preventivas a fim de reduzir a taxa de inadimplência institucional.

É importante ressaltar também, como uma das possíveis consequências do aumento da inadimplência nas IES nos últimos anos, o crescimento de 182,5% entre 2008 e 2018 do número de matrículas em cursos de educação a distância (EAD), enquanto na modalidade presencial o crescimento foi apenas de 25,9% nesse mesmo período (INEP, 2018). Muitos alunos, com dificuldade em se manterem financeiramente em um curso de graduação presencial, optam por migrar para cursos a distância, que oferecem valores de mensalidades mais baixos, além de permitirem maior flexibilidade de horário e local de estudo. De acordo com o SEMERJ (2019), em 2019, no Estado do Rio de Janeiro, o valor médio das mensalidades em cursos presenciais no primeiro semestre ficou em torno de R\$ 1,4 mil, enquanto na modalidade a distância o valor foi de aproximadamente R\$ 428.

A consequência de maior impacto da inadimplência para as IES, além das questões financeiras, é a evasão estudantil. Alunos com dificuldades em arcar com os valores das mensalidades, acabam por abandonar seus cursos. Em 2018, a taxa de evasão anual nos cursos presenciais do Rio de Janeiro chegou a 31,7%, sendo 34,5% na rede privada e 22,8% na rede pública (SEMERJ, 2019).

A evasão estudantil é um problema internacional que afeta o resultado dos sistemas educacionais. As perdas de estudantes que iniciam, mas não terminam seus cursos são desperdícios sociais, acadêmicos e econômicos (SILVA FILHO, 2007). Do ponto de vista econômico, a evasão representa uma perda de receita para a instituição, e um maior esforço de captação para manter o nível da base de alunos. Além disso, uma maior taxa de evasão influencia o valor de um aluno para a instituição, que representa não só o que este aluno irá pagar ao longo de todo o curso, mas também o potencial dele continuar sendo um cliente no futuro, procurando por cursos de pós-graduação e extensão. Uma base de alunos com menor taxa de evasão representa uma base de clientes mais valiosa para a instituição (PEREIRA, 2017). Ademais, a evasão estudantil representa também um custo social para as instituições, uma vez que é necessário um determinado número de

concluintes no ensino superior para se atingir a meta do Plano Nacional de Educação (PNE) em 2024, e, quando um aluno evade, ele representa menos um aluno formado que poderia estar contribuindo para o desenvolvimento socioeconômico do país.

O uso de modelos de *credit scoring*, apesar de bastante difundido em instituições financeiras, é pouco utilizado no setor de educação superior privada. O *credit scoring* é um dos principais modelos quantitativos de análise de risco de crédito. Trata-se de um conjunto de modelos de decisão e técnicas estatísticas que ajudam os emprestadores de crédito a discriminar entre os bons e maus clientes, a identificar o quanto de crédito conceder e a estabelecer qual estratégia operacional irá aumentar sua rentabilidade com cada cliente (THOMAS, 2000). De acordo com Santos e Famá (2007), os modelos de *credit scoring* amparam-se em pontuações estatísticas de risco, um instrumento desenvolvido para avaliar a probabilidade de que determinado cliente venha a tornar-se inadimplente no futuro, permitindo a classificação destes em grupos distintos, de acordo com sua pontuação.

As poucas IES que realizam alguma atividade no sentido de avaliação do risco de crédito limitam-se apenas a consultas no Serviço de Proteção ao Crédito (SPC) e ao Serasa Experian, sem que, no entanto, haja uma real intenção de limitar o acesso aos serviços da empresa, ou diferenciar os alunos de acordo com seu potencial risco de se tornar inadimplente. Além disso, segundo a legislação educacional vigente, a relação existente entre a IES e seus alunos e responsáveis é considerada uma relação de consumo e, como determina o art. 6º da Lei 9.870 de 23 de novembro de 1999, o aluno não pode sofrer qualquer penalidade pedagógica ou constrangimento pela falta de pagamento. As IES não podem impedir o aluno de frequentar as aulas, fazer provas ou reter documentos escolares, entretanto, a lei permite que as IES não sejam obrigadas a renovar a matrícula do aluno devedor para o período letivo subsequente. Com isso, qualquer ação tomada pelas IES contra a inadimplência é dificultada pela questão legal supracitada, uma vez que só é possível limitar o acesso dos alunos aos serviços educacionais ao final de cada semestre, e não no momento em que ele se torna inadimplente.

Assim exposto, as IES, em sua maioria, não utilizam qualquer tipo de técnica de *credit scoring* para diferenciar seus alunos quanto ao risco de se tornarem inadimplentes, falhando em entender também quais fatores são mais relevantes para esse comportamento, e, com isso, não sendo capazes de planejar políticas

preventivas de natureza pedagógica, acadêmica e financeira específicas para cada aluno. De acordo com Bem et al. (2008), existem dois fatores determinantes do risco de inadimplência em uma IES: a fraca qualidade no processo de análise de crédito (fator interno) e o agravamento da situação macroeconômica, que pode resultar na escassez de tomadores saudáveis (fator externo). Dado o aspecto incontrolável dos fatores externos, é importante para as IES buscarem um processo de análise de crédito que consiga prever, com a maior exatidão possível, o risco de inadimplência de cada aluno, para que seja possível a adoção de medidas preventivas.

Entende-se, assim, a importância do uso de técnicas de mineração de dados e *credit scoring* para auxiliar na descoberta dos fatores que determinam a capacidade pagadora da mensalidade de estudantes e, conseqüentemente, as principais causas da inadimplência estudantil em uma IES privada. Assim, será possível realizar ações preventivas individualizadas para identificar alunos com potencial risco de se tornarem inadimplentes, bem como premiar os adimplentes com eventuais descontos em mensalidades, reduzindo, também, a taxa de evasão institucional.

1.2. Objetivo do estudo

O objetivo deste trabalho é propor e testar um modelo de *credit scoring* com o uso de técnicas de mineração de dados no setor de educação superior privada, utilizando dados financeiros, acadêmicos e sociais de alunos de uma IES privada do Rio de Janeiro, verificando os fatores que mais influenciam este comportamento e estimando o risco de inadimplência de cada aluno.

1.3. Relevância do estudo

As organizações devem gerenciar seu relacionamento com os clientes de forma eficaz para se manterem competitivas. Os avanços tecnológicos permitiram que, por meio do uso de técnicas de *Customer Relationship Management* (CRM), as empresas gerenciassem o relacionamento com seus clientes de maneira mais eficiente, mas a tecnologia também permitiu que os clientes se informassem e exigissem muito mais das empresas com as quais fazem negócios (PEPPERS; ROGERS, 2011). As empresas dependem, cada vez mais, da grande quantidade de

dados sobre clientes para poder identificá-los e diferenciá-los, oferecendo produtos e serviços cada vez mais personalizados e de acordo as necessidades de cada cliente.

O gerenciamento do risco de crédito ao consumidor é uma questão importante quando se trata de empresas que dependem do planejamento financeiro de seus recebíveis para gerenciar suas operações (SANTOS; FAMÁ, 2007). Para essas empresas, saber quais consumidores merecem crédito significa poder para reduzir custos operacionais, reduzir dívidas incobráveis e obter maior lucro.

De acordo com Lemos, Ribeiro e Siqueira (2017), a importância de avaliar o comportamento de clientes antigos e futuros consumidores em termos de valor do crédito é crucial. Da mesma forma, a capacidade de discernir entre candidatos bons e ruins o mais cedo possível, de preferência durante a fase inicial do empréstimo, é inestimável. Todo o risco nos negócios advém de não saber de antemão quais candidatos serão consumidores rentáveis e quais serão inadimplentes. As perdas podem advir da aceitação de um candidato ruim ou da recusa de um candidato bom, à medida que atrair novos clientes lucrativos e manter bons e antigos exige que as empresas sejam mais eficientes e eficazes. É neste momento que se apresenta a relevância da análise de risco de crédito ao consumidor (CASA NOVA, 2013).

Os últimos anos trouxeram grandes avanços nas ciências da computação e estatística. O julgamento humano, quase sempre subjetivo e baseado em poucos dados, foi substituído por modelos estatísticos poderosos e robustos, alimentados por bases de dados ricas em informação. Esforçando-se para obter melhores resultados e fazer uso mais inteligente dos bancos de dados de consumidores disponíveis, as empresas de crédito buscam soluções analíticas que possam ajudá-las a ampliar seu conhecimento do consumidor e a tomar as melhores decisões possíveis a respeito de sua estratégia competitiva (GONÇALVES; GOUVÊA; MANTOVANI, 2013).

No setor de educação, o CRM tem sido empregado aquém de sua capacidade, somente como uma solução tecnológica de *software* para auxiliar no envio de *e-mail* marketing e ligações da área de *call center*, e não como um aliado estratégico para as empresas que almejam ter foco no cliente (PEPPERS; ROGERS, 2011). Os gestores justificam o mau uso do CRM pelo fato de ser difícil diferenciar os clientes no setor de educação superior, uma vez que os valores de mensalidades são os mesmos, exceto em casos de bolsa de estudos. Entretanto, por que não diferenciar os alunos entre aqueles com maior risco de se tornarem inadimplentes,

entendendo os fatores com maior influência nesta questão? Além de auxiliar na redução da taxa de inadimplência, trará impactos positivos também para os índices de evasão das IES.

A maior parte dos estudos da literatura no Brasil acerca do tema da inadimplência são voltados para o setor financeiro (ARAÚJO E CARMONA, 2007; SANTOS E FAMÁ, 2007; ANNIBAL et al., 2009; CAMARGOS et al., 2010; ALVES E CAMARGOS, 2014). Sobrinho (2017), apresenta um estudo sobre o uso de modelos de *credit scoring* na educação básica e Herling et al. (2013), Freire e Freire (2014) e Lemos, Ribeiro e Siqueira (2017) estudam a inadimplência em IES, no entanto, de forma mais qualitativa, com pouco aprofundamento estatístico pela falta de facilidade de acesso aos dados. Assim, a relevância deste trabalho está em usufruir do acesso aos dados de alunos de uma IES privada e propor algo inédito no setor de educação superior, que é o uso de modelos de *credit scoring* como um instrumento de gestão estratégica para as IES, para que elas sejam capazes de diferenciar seus alunos de acordo com seu respectivo risco de se tornar inadimplente e identificar fatores de natureza social, financeira e acadêmica que tenham maior relação com este risco. Dessa forma, as IES poderão realizar ações preventivas e customizar não só o produto oferecido (no caso, os cursos de graduação), mas também todo o pacote de serviços e esforços envolvidos, além de conseguirem planejar suas ações de comunicação de retenção voltadas para uma estratégia de valor do cliente – risco de inadimplência.

1.4. Delimitação do estudo

A proposta deste estudo é analisar exclusivamente o comportamento de um grupo de alunos de uma única IES privada do Rio de Janeiro e estabelecer relações entre variáveis financeiras, acadêmicas e sociais e o possível risco de inadimplência. A base teórica a ser abordada no estudo será pautada pelos conceitos de CRM, *credit scoring* e inadimplência em IES.

Serão utilizadas informações existentes no banco de dados de uma IES privada do Rio de Janeiro e analisados dados de aproximadamente 5 mil alunos que estiveram, durante os anos de 2017 e 2018, adimplentes ou inadimplentes em uma ou mais mensalidades de seu curso. Para a IES avaliada, inadimplente é aquele

aluno que tem de uma a cinco mensalidades em aberto. A partir do quinto dia após a data de vencimento do boleto de cobrança, o aluno é considerado inadimplente.

Alunos que não estejam devendo nenhuma mensalidade, porém recorrentemente realizam o pagamento fora da data de vencimento, são classificados pela IES avaliada como imptuais, e serão considerados, neste estudo, como adimplentes. Vale ressaltar também que, dada a natureza do serviço prestado e a impossibilidade de limitar o acesso aos serviços educacionais no momento em que o cliente/aluno se torna inadimplente, a condição de inadimplência será analisada, neste estudo, ao final de cada semestre.

1.5. Estrutura do trabalho

Após o capítulo introdutório, os capítulos a seguir estão organizados da seguinte maneira: o capítulo 2 apresenta o referencial teórico do estudo, que versa acerca da temática de *credit scoring* e inadimplência no setor de educação superior. Já o capítulo 3 dedica-se a explicar os procedimentos metodológicos da pesquisa. Em seguida, o capítulo 4 analisa os dados coletados e, por fim, o capítulo 5 apresenta as principais conclusões alcançadas.

2 Referencial Teórico

Este estudo pretende abordar a inadimplência no ensino superior privado como um problema de gestão de marketing. Os conceitos teóricos utilizados baseiam-se em CRM, *credit scoring*, inadimplência e técnicas de mineração de dados, e têm como função identificar o risco de inadimplência dos alunos para que eles possam ser tratados de forma individualizada, aumentando a eficácia das ações de retenção de marketing, além de proporcionar a IES uma redução de custos operacionais e um planejamento financeiro mais sadio.

2.1. Customer Relationship Management

De acordo com Peppers e Rogers (2011), CRM é uma estratégia de negócios que envolve toda a empresa, focada em gerenciar o relacionamento com o cliente. Os autores destacam também que, de forma geral, criar valor para uma base de clientes significa tratar diferentes clientes de forma diferenciada, e o foco no cliente é uma estratégia cada vez mais usada no ambiente empresarial, uma vez que todos os clientes, em todas as fases de vida e em todas as indústrias estão demandando atendimento individualizado e personalizado. Além disso, trata-se de uma forma mais eficiente de fazer negócios.

Para aumentar o valor de uma base de clientes, as empresas precisam atentar para a tríade *Get-Keep-Grow* (PEPPERS E ROGERS, 2011). *Get* significa adquirir novos clientes lucrativos para a empresa, *Keep* está relacionado a retenção dos clientes lucrativos e eliminação daqueles que não trazem benefícios financeiros para a empresa, e *Grow* ao aumento do número de serviços/produtos adquiridos por um cliente. No caso deste trabalho, será usada principalmente a perspectiva *Keep*, no sentido de entender o risco de cada aluno tornar-se inadimplente e possivelmente vir a evadir, e, assim, poder realizar ações para retê-lo. A perspectiva *Get* também poderá ser utilizada no intuito da IES procurar captar principalmente alunos que tenham menor risco de se tornarem inadimplentes no futuro, e a *Grow* no sentido

de aumentar o valor da base de clientes, oferecendo cursos de extensão e pós-graduação para os egressos que apresentaram um histórico de adimplência.

Kumar (2018) propõe uma Teoria de Avaliação dos Clientes (CVT) baseada em princípios econômicos e que conceitualiza a geração de valor dos clientes para as empresas. Segundo o autor, qualquer negócio sustentável cria valor para os clientes através da oferta de produtos e serviços, e extrai valor deles através do lucro com as vendas. Entretanto, para Kumar et al. (2010), avaliar o valor de um cliente levando em consideração somente o aspecto transacional pode não ser suficiente, uma vez que, atualmente, os clientes se relacionam com a empresa de diversas formas e por diversos canais, de modo que a criação de um relacionamento de confiança e duradouro é fundamental para garantir o engajamento do cliente com a empresa. Ainda de acordo com o autor supracitado, o valor de um cliente para uma empresa pode ser medido por meio de aspectos sociais, como o engajamento nas redes sociais da empresa, propensão a indicar novos clientes, conhecimento sobre os produtos e serviços oferecidos pela empresa, além dos aspectos financeiros.

No setor de educação superior, o conceito de valor do aluno é pouco utilizado. O aluno não é diferenciado de acordo com suas necessidades e seu valor para a instituição, de forma que receba um tratamento individualizado. As IES privadas selecionam os alunos exclusivamente com base na nota do Exame Nacional do Ensino Médio (ENEM), ou pela capacidade de pagamento do aluno, no caso de escolas de elite. As IES públicas realizam uma seleção pautada em aspectos étnicos e socioeconômicos. Após ingressarem nas instituições, até mesmo a diferenciação em termos de histórico de transações (financeira) não é realizada, pela percepção de que todos os alunos pagam o mesmo valor de mensalidade, exceto aqueles com bolsas de estudo, e, assim, todos os alunos teriam o mesmo valor para as IES.

Há evidências empíricas na literatura (PEPPERS E ROGERS, 2011) da relação entre o nível de utilização de ferramentas de CRM com a lealdade de longo prazo do cliente com a empresa. O cliente reconhece valor no tratamento individualizado que ele recebe. Segundo Barnes (2001), o aumento do valor percebido pelo cliente em cada interação com a empresa aumenta seu nível de satisfação, levando a taxas maiores de retenção. É interesse ressaltar, também, que o conceito de valor do cliente pode ser diferente de uma empresa para a outra (PEPPERS E ROGERS, 2011). Sendo assim, entender quais atributos de um aluno

são mais relevantes para uma IES e diferenciá-lo, seja de acordo com suas transações financeiras, o seu potencial risco de se tornar inadimplente, ou de evadir, ou ainda de acordo com o seu poder de influência para trazer novos alunos ou, até mesmo, uma diferenciação por desempenho acadêmico, levando a IES a melhores resultados no Exame Nacional de Desempenho dos Estudantes (ENADE), são de extrema importância para as instituições que desejam realizar uma gestão baseada nos princípios de CRM.

2.2. Gerenciamento de crédito ao consumidor

Crédito é “todo ato de vontade ou disposição de alguém de destacar ou ceder, temporariamente, parte do seu patrimônio a um terceiro, com a expectativa de que essa parcela volte a sua posse integralmente, após decorrido o tempo estipulado” (SCHRICKEL, 1998, p. 25). Para Matias (2007), o crédito está associado à troca de bens e serviços oferecidos no tempo presente com a expectativa de ser remunerado futuramente. Camargos et al. (2010) acrescentam ainda a questão dos juros como remuneração pela cessão temporária de recursos a terceiros e Santos e Famá (2007) ressaltam a importância do crédito como uma fonte de financiamento às pessoas físicas, principalmente em situações de descasamento dos prazos de recebimento da renda e pagamento das despesas.

A concessão de crédito a terceiros é uma questão relevante para as grandes indústrias que operam no mundo, facilitando as transações de bens ou serviços e permitindo que os varejistas impulsionem suas vendas (GONÇALVES; GOUVÊA; MANTOVANI, 2013). Segundo Brigham, Gapenski e Ehrhardt (2001, p. 794), “a oferta de crédito por parte de empresas e instituições financeiras é um importante impulsionador da atividade econômica, por disponibilizar recursos financeiros às pessoas físicas para que possam financiar suas necessidades permanentes e eventuais”. De maneira mais ampla, Palmuti e Picchiali (2012) definem o crédito como um instrumento de desenvolvimento da economia, uma vez que permite o financiamento de agentes econômicos, tais como o Estado, as empresas e as famílias, para que possam satisfazer suas necessidades de consumo e investimento.

Sendo assim, as empresas precisam adaptar suas políticas de crédito de acordo com suas capacidades financeiras e demandas do mercado em que estão inseridas (HERLING et al., 2013). De acordo com Ross (2015), a política de crédito

influencia diretamente o modo como as empresas irão conduzir seus negócios e pode ser definida por três aspectos: as condições de venda, seja ela à vista, ou a prazo; uma análise de crédito que, através de técnicas estatísticas, permita distinguir entre clientes que pagarão as contas e clientes que não o farão e uma política de cobrança bem estabelecida, processo fundamental para empresas que concedem crédito a terceiros.

Diante do panorama de crise econômica que o Brasil e particularmente o estado do Rio de Janeiro se encontram nos últimos anos, a questão da concessão de crédito a terceiros está cada vez mais recorrente em todos os setores de negócio, e não exclusivamente em instituições financeiras. No setor de educação superior privada, com a redução, nos últimos anos, do número de vagas ofertadas pelo FIES, as IES passaram a criar programas próprios de financiamento e concessão de crédito estudantil. De acordo com Lemos, Ribeiro e Siqueira (2017), nos programas próprios de financiamento, as IES tornam-se responsáveis por todo o processo de concessão de crédito, desde a análise de risco de crédito até a implementação do empréstimo, incluindo o recebimento dos valores e posteriores cobranças. Entretanto, recai também sobre as IES todo o risco de inadimplência do não pagamento, e, dessa forma, fica clara a importância de realizarem uma análise de concessão de crédito que permita prever com a maior exatidão possível o potencial risco de cada aluno vir a se tornar inadimplente, para que sejam adotadas medidas preventivas.

2.3. Avaliações do risco de crédito

Gitman (2001) define o risco como sendo a possibilidade de um prejuízo financeiro, em que os ativos que possuem maiores probabilidades de prejuízos são vistos como os mais arriscados em detrimento daqueles com menores probabilidades. Para esse autor, o termo risco é usado alternativamente como incerteza, ao referir-se à variabilidade de retornos associada a um ativo. Ainda de acordo com Gitman (2001), diante da possibilidade de perda é que as instituições, sejam elas financeiras ou não, sentem, cada vez mais, a necessidade de ter acesso a mecanismos que as possibilitem um gerenciamento ativo de seus riscos.

Para Lemos, Ribeiro e Siqueira (2017), o risco é algo inerente a todo o tipo de negócio, e como em qualquer outro, a educação superior no Brasil não está livre

de ameaças ambientais, como concorrência, evasão estudantil, inadimplência, entre outros, que reduzem a capacidade competitiva das IES particulares.

O risco da concessão de crédito a terceiros, especificamente, advém do fato do crédito se referir à troca de um valor presente por uma promessa de reembolso futuro, não necessariamente certa, em virtude do “fator risco”. Por envolver a expectativa do retorno do patrimônio, deve-se entender que todo crédito está associado a um certo risco (GONÇALVES; GOUVÊA; MANTOVANI, 2013).

Santos (2003) ressalta ainda a importância de considerar os fatores externos, ou sistemáticos, na determinação do risco de crédito. A volatilidade das taxas de juros, recessão econômica e nível de desemprego, por exemplo, podem afetar a capacidade pagadora dos tomadores. E, dado o aspecto incontrolável dos fatores externos, é de suma importância para as organizações que pretendem realizar análise de crédito de seus clientes, focarem na qualidade do processo. Todo o risco do negócio advém de não saber de antemão quais candidatos serão consumidores rentáveis e quais serão inadimplentes e, assim, acabar aceitando um candidato ruim ou recusando um candidato bom (CASA NOVA, 2013).

Como forma de minimizar o risco de inadimplência, Santos e Famá (2007) destacam a importância da gestão do risco de crédito, baseada em procedimentos subjetivos (análise caso a caso ou sistemas especialistas) e objetivos (análise estatística), como instrumento para a adequada seleção, análise, precificação e, principalmente, monitoramento do risco de inadimplência, quando da ocorrência de fatores sistêmicos adversos. Para Gonçalves, Gouvêa e Mantovani (2013), a avaliação subjetiva do risco de um potencial cliente pode ser feita por meio de julgamento, que envolve uma análise mais qualitativa, enquanto uma avaliação mais objetiva e quantitativa deve ser realizada por meio da classificação do tomador, via modelos de avaliação, como os de *credit scoring*.

2.3.1. Avaliações subjetivas de crédito

De acordo com Santos (2003, p. 46), “a análise subjetiva de crédito é baseada na experiência adquirida dos analistas de crédito, no conhecimento técnico, no bom-senso e na disponibilidade de informações (internas e externas) que lhes possibilitem diagnosticar se o cliente possui idoneidade e capacidade de gerar receita para honrar o pagamento das parcelas dos financiamentos”. Ainda segundo

o autor, estas avaliações podem ser entendidas como uma metodologia de análise de crédito que utiliza avaliações subjetivas para identificar a capacidade do solicitante de financiamento em arcar com as despesas do crédito.

Os analistas de crédito utilizam frequentemente os 5 Cs do crédito (caráter, capacidade, capital, colateral e condições), largamente difundido na literatura (GITMAN, 2001; ROSS et al., 2015), como importantes condutores de valor para a decisão de concessão de crédito, tanto para pessoas físicas, como jurídicas. Através de informações como idoneidade (caráter), habilidade em converter negócios em renda (capacidade), situação financeira e capacidade de pagamento (capital), disponibilidade de bens móveis, imóveis e financeiros (colateral) e impacto dos fatores externos na geração de fluxos de caixas (condições), as decisões de crédito são baseadas e ponderadas de acordo com os modelos desenvolvidos para a previsão de inadimplência, utilizados na análise objetiva (SANTOS E FAMÁ, 2007).

Alguns autores consideram ainda um sexto critério de avaliação subjetiva de risco de crédito, o conglomerado, que diz respeito a análise do grupo de empresas no qual o solicitante do crédito está incluído. Segundo Sobrinho (2007), o intuito deste tipo de análise é ter um panorama da solidez do grupo de empresas que, de certa forma, estão interligadas entre si e a possibilidade de realizar uma avaliação mais completa, uma vez que o mau desempenho de algumas das empresas do grupo, pode afetar a solidez das demais. Além disso, Blatt (1999) ressalta que se deve estender também a análise para todos os indivíduos que possivelmente possam responder solidariamente a um débito, como cônjuges, dependentes, garantidores, entre outros.

Apesar da grande importância das avaliações subjetivas, existem algumas limitações relacionadas ao seu uso. A primeira delas diz respeito ao fato de tratar-se de um sistema dispendioso, uma vez que necessita de uma grande quantidade de analistas para fazer frente a demanda de recursos, além de exigir treinamentos contínuos a fim de uniformizar o padrão de análise. Outro ponto de atenção com relação as análises subjetivas de crédito é que, se utilizadas de forma isolada, sem fundamentos estatísticos complementares, tornam sua aplicação frágil, uma vez que dificultam a quantificação do risco de crédito, inviabilizando a identificação dos fatores que estatisticamente influenciam a inadimplência. Além disso, a natureza

subjetiva da avaliação torna sua interpretação sujeita a influência de valores individuais dos analistas (SOBRINHO, 2007).

2.3.2. Avaliações objetivas de crédito – *credit scoring*

As avaliações objetivas do risco baseiam-se em metodologias estatísticas, com a finalidade de apurar resultados matemáticos que atestem a capacidade de pagamento dos tomadores. Ampara-se em pontuações estatísticas de risco, um instrumento desenvolvido para que o analista avalie a probabilidade de que determinado cliente venha a tornar-se inadimplente no futuro (SANTOS E FAMÁ, 2007).

Casa Nova (2013) destaca a importância de se utilizarem modelos estruturados e com bom ajuste para a análise de crédito, pois reduzem o risco de se conceder crédito a potenciais maus pagadores. Além disso, segundo o autor, há um custo mais elevado em se classificar erroneamente um mau pagador do que um bom pagador, isto é, há maior dano à organização se um mau cliente for classificado como bom pagador do que a situação contrária. Dessa forma, é necessário observar com cuidado a taxa de acertos de classificação dos inadimplentes.

Dentre as técnicas objetivas de gestão do risco de crédito, destaca-se o *credit scoring*. De acordo com Thomas (2000), os modelos de *credit scoring* são sistemas que atribuem pontuações às variáveis de decisão de crédito de um proponente, mediante a aplicação de técnicas estatísticas. Esses modelos visam a segregação de características que permitam distinguir os bons dos maus créditos. Ainda segundo o autor, os modelos de *credit scoring* podem ser percebidos, também, como mecanismos de status entre diferentes grupos, uma vez que ao realizar uma previsão de um evento, classifica os indivíduos entre aqueles adimplentes e inadimplentes.

A partir de uma equação gerada por meio de variáveis referentes ao proponente ou à operação de crédito, os sistemas de *credit scoring* geram uma pontuação que representa o risco de perda. Esta pontuação, ou *score*, pode ser interpretada como a probabilidade de inadimplência ao se comparar com a pontuação estabelecida como ponto de corte ou pontuação mínima aceitável, que servirá como base para a aprovação ou recusa do crédito (ARAÚJO E CARMONA, 2007).

O ponto de corte deve ser estabelecido de modo que otimize a relação entre a aceitação de bons clientes e rejeite o máximo possível a aceitação de inadimplentes (SOBRINHO, 2007). Os erros relacionados a rejeição de bons clientes (erro tipo I) e a aceitação de maus clientes (erro tipo II) levam a empresa credora à perdas e, portanto, as políticas de ponto de corte das empresas devem ser bem definidas a fim de que as perdas incorridas sejam compensadas pela identificação de bons clientes (MCCLAVE et al., 2014).

Os modelos de *credit scoring* podem ser divididos em duas categorias, quanto aos seus objetivos: modelos de aprovação de crédito e modelos de escoragem comportamental, também conhecidos por *behavioural scoring* (SAUNDERS, 2000). Os modelos de aprovação de crédito utilizam dados sobre a vida pessoal e profissional do solicitante de crédito para prever o comportamento futuro dos atuais e novos clientes, em relação ao pagamento de débitos nos termos do acordo de crédito (ROSENBERG E GLEIT, 1994). Thomas (2000), acrescenta ainda que estes modelos são ferramentas que dão suporte à tomada de decisão sobre a concessão de crédito para novas aplicações ou novos clientes.

Já os modelos de *behavioural scoring* também visam a previsão da inadimplência, porém o seu foco é na análise de indivíduos que já são clientes da instituição e possuem uma relação creditícia, possibilitando a incorporação do histórico de pagamento destes clientes ao conjunto de informações preditoras do modelo (SOBRINHO, 2007). Para Thomas (2000), ao absorver estas informações passadas, o *behavioural scoring* pode ser percebido como um mecanismo que possibilita o gerenciamento de créditos existentes, o que permite a tomada de decisão em relação as alterações de limite de crédito, realização de cobranças preventivas, administração da recuperação de débitos, entre outros.

2.4. Aplicações de *credit scoring* em diferentes setores de negócio

No setor de educação superior privada, o uso de modelos de *credit scoring* encontra no problema de inadimplência estudantil uma oportunidade eficaz de aplicação. Os dados gerados usualmente sobre alunos, relacionados a desempenho acadêmico atual e progresso, informações demográficas coletadas no ato da matrícula e informações do relacionamento financeiro com a instituição podem indicar padrões associados a capacidade de pagamento. Há casos publicados na

academia (SOBRINHO, 2007; SANTOS E FAMÁ, 2007; AKKOÇ, 2012) que ilustram a relação positiva entre o uso de modelos de *credit scoring* e o controle do risco de inadimplência, principalmente no setor financeiro, onde o uso destes modelos é mais difundido. A seguir, serão resumidos três trabalhos acadêmicos, em que são aplicadas metodologias similares.

2.4.1. Aplicação de *credit scoring* no setor de educação básica

Sobrinho (2007) desenvolveu um estudo em uma instituição de ensino voltada para o ensino médio e fundamental em Recife, com o intuito de verificar a viabilidade da utilização de modelos de *credit scoring*, como uma ferramenta de previsão de risco, no controle da inadimplência.

Neste estudo, a viabilidade do *credit scoring* foi testada através da análise discriminante, regressão logística e redes neurais artificiais. Os dados utilizados no estudo foram retirados de uma base de 303 alunos, dos quais 245 foram considerados aptos para serem analisados. Eles referem-se às características dos alunos, pais e responsáveis financeiros que usufruem dos serviços educacionais prestados pela instituição avaliada.

Os dados foram analisados através de duas categorizações dos modelos de *credit scoring*, os modelos de aprovação de crédito e os modelos de *behavioural scoring*. Com relação aos resultados, a técnica de análise discriminante classificou corretamente 72.13% dos indivíduos no modelo de aprovação de crédito e 81.97% no *behavioral scoring*. Na regressão logística, as observações corretamente classificadas alcançaram 80.33% no modelo de aprovação de crédito e 88.52% no *behavioral scoring*. Nas redes neurais, o modelo de aprovação de crédito alcançou 86.89% de acerto em sua previsão, enquanto o *behavioral scoring* atingiu o patamar de 88.52% de casos identificados corretamente.

De acordo com Sobrinho (2007), pelos resultados obtidos, é possível afirmar que todos os modelos elaborados se apresentaram como instrumentos viáveis na previsão de inadimplência, uma vez que seu poder de previsão, quanto ao status de pagamento de clientes, é superior ao critério de identificação baseado em chances proporcionais, o qual permite um acerto mínimo de 62.5%. Ainda de acordo com o autor, como conclusão do estudo, pode-se afirmar que os modelos de *credit scoring* indicaram ser de grande valia à medida que se apresentaram como

instrumento de previsão viável, permitindo sua utilização como instrumento auxiliar na seleção de potenciais clientes ou ainda como ferramenta de apoio ao planejamento financeiro.

2.4.2. Aplicação de *credit scoring* no setor de crédito bancário de pessoas físicas

A pesquisa realizada por Santos e Famá (2007) tem como objetivo propor a utilização de um modelo de *credit scoring* para créditos rotativos composto por variáveis sistêmicas e não-sistêmicas diretamente relacionadas à capacidade de pagamento de pessoas físicas com renda assalariada, e verificar se a utilização do modelo contribuiria para a redução do atual risco de inadimplência em créditos rotativos.

O modelo de *credit scoring* aplicou a fórmula em uma amostra de 2.000 propostas de crédito rotativo de pessoas físicas com renda assalariada que apresentavam situação ativa no Banco X, ou seja, estavam utilizando seus limites de crédito. Os 2.000 clientes ativos apresentavam situação de baixo e elevado risco em utilizações de créditos rotativos no momento da testagem do modelo, sendo distribuídos em duas categorias: 1.000 “clientes prospectivos” com favorável histórico de crédito e índice de comprometimento da renda assalariada de até 30% com o pagamento de juros e parcelas em créditos rotativos; os restantes 1.000 foram classificados em “clientes não-prospectivos” por apresentarem elevada utilização de créditos rotativos, índice de comprometimento da renda assalariada superior a 30% com o pagamento de juros e parcelas em créditos rotativos e registro de informações restritivas no mercado. O objetivo foi verificar se o modelo identificaria as situações correntes de risco de cada cliente através de suas notas confrontadas com o ponto de corte. O ponto de corte foi de 360 pontos, assumindo-se um intervalo de confiança de 5%, ou seja, para pontuações entre 342 e 378 a decisão de crédito se daria mediante análise colegiada em comitê.

Os resultados extraídos da amostra mostraram que do total de 1.000 clientes na condição inicial de “prospectivos”, 892 foram reclassificados na posição de “não-prospectivos”, ou seja, não teriam suas propostas de crédito aceitas, 48 clientes seriam encaminhados para análise colegiada e os restantes 60 receberiam aprovação. Verifica-se, assim, que o modelo contribuiria com uma significativa

redução de risco de 89,20%, que poderia chegar a até 94,00%, caso a análise colegiada também desse parecer desfavorável às propostas de crédito.

Sendo assim, de acordo com Santos e Famá (2007), constatou-se que o modelo de *credit scoring* proposto capturaria melhor a situação corrente de risco do Banco X por incluir um conjunto maior de variáveis representativas da real situação de risco dos clientes (total de 23 variáveis), e por ponderá-las de acordo com o seu justo ou aproximado grau de importância. Além disso, de acordo com os autores, o modelo atual utilizado pelo Banco X não considerava o impacto de fatores sistêmicos na renda e capacidade de pagamento dos tomadores, ele incluía um conjunto de apenas 10 variáveis de natureza demográfica, financeira, patrimonial e de idoneidade dos clientes.

2.4.3. Aplicação de *credit scoring* no setor de cartões de crédito

O estudo realizado por Akkoç (2012) propõe um modelo híbrido de *credit scoring* do Sistema de Inferência Neuro Fuzzy Adaptativo (ANFIS), baseado em técnicas estatísticas e Neuro Fuzzy. O desempenho do modelo proposto foi comparado com modelos convencionais e comumente atualizados. Os modelos de *credit scoring* foram testados usando um processo de validação cruzada de 10 vezes (*10-fold cross-validation*) com os dados do cartão de crédito de um banco internacional que opera na Turquia.

O ANFIS, desenvolvido por Yang (1993), é um tipo de sistema da Neuro Fuzzy que utiliza a experiência humana na forma de regras “*if-then*” e tem a capacidade de construir modelos com dados de amostra de destino. De acordo com Jang et al. (1997), o ANFIS determina a si próprio parâmetros apropriados para fornecer o melhor aprendizado, descrevendo as funções de associação e inserindo dados no sistema.

Neste estudo, os modelos de *credit scoring* são desenvolvidos usando dados de cartão de crédito de um banco internacional que opera na Turquia. Os dados consistem em um conjunto de empréstimos concedidos em 2003 a um total de 2.000 solicitantes de cartão de crédito, dos quais 1.000 são considerados bons consumidores de crédito, enquanto os outros 1.000 são considerados maus consumidores de crédito. Neste caso, um consumidor de crédito ruim é definido como aquele que está inadimplente em, pelo menos, três pagamentos no período de

um ano, caso contrário, é definido como um bom consumidor de crédito. Para minimizar o possível viés associado à amostragem aleatória, utilizou-se a validação cruzada de 10 vezes.

Com relação aos resultados, nos modelos de *credit scoring* que utilizam técnicas de análise discriminante linear, a taxa correta média de classificação é de 57,2% e o modelo classificou corretamente 49,4% dos bons solicitantes de cartão de crédito e 65% dos maus solicitantes de cartão de crédito. Nos modelos que utilizam a regressão logística, a taxa média de classificação é de 57,75%, classificando corretamente 50,54% dos solicitantes de cartão de crédito bons e 65% dos maus solicitantes. Nos modelos de redes neurais, a taxa média correta de classificação é de 58,6% e o modelo classificou corretamente 75% dos solicitantes de bons cartões de crédito e 42,2% dos maus solicitantes de cartões de crédito. Por fim, no modelo proposto pelo estudo, a taxa média de classificação correta é de 60%, classificando corretamente 51,4% dos bons solicitantes de cartão de crédito e 68,6% dos maus solicitantes de cartão de crédito.

Para Akkoç (2012), os resultados indicaram que o modelo híbrido proposto apresenta desempenho melhor do que as abordagens de análise discriminante linear, análise de regressão logística e rede neural artificial (RNA) em termos de taxa de classificação correta média e custo estimado de classificação. Como na RNA, o modelo proposto possui capacidade de aprendizado, e diferente dela, o modelo não fica em uma caixa preta. No modelo proposto, a interpretação de variáveis independentes pode fornecer informações valiosas para banqueiros e consumidores, especialmente na explicação de por que os pedidos de crédito são rejeitados.

2.5. Inadimplência no setor de educação superior privada

A inadimplência é definida por Teixeira e Silva (2001, p.19) “como a falta de pagamento; inadimplemento é o termo jurídico utilizado, em regra, para designar uma situação de não cumprimento da cláusula contratual; insolvência é a perda total de capacidade de pagamento”. De acordo com Herling et al. (2013), um indivíduo torna-se propício a ser inadimplente à medida que sofre constantemente impactos na sua renda e esses impactos determinam os seus limites e restrições orçamentárias, fazendo com que ele perca seu controle financeiro. Para Maia (2007) a inadimplência tem um impacto econômico e social negativo, já que atrasa o

retorno do capital ao mercado, encarece o acesso ao crédito, além de retardar a expansão do negócio e conseqüentemente a geração de emprego.

Teixeira e Silva (2001) definem como “devedor crônico” aquele que sempre atrasa o pagamento, mas acaba pagando. Para os autores, o “verdadeiro mau pagador” é definido como uma pessoa com a intenção de lesar o credor, e recusa pagar o débito ou tenta prolongar ao máximo o pagamento. Já o “mau pagador ocasional” não tem a intenção de enganar o credor, porém por motivos pessoais não tem condições de honrar seus compromissos.

De acordo com Herling et al. (2013), diante da atual situação econômica brasileira, muitos alunos passaram a ser o “mau pagador ocasional”, dado que com a situação financeira abalada, os alunos priorizam pagamentos com bens de consumo, considerados essenciais à sobrevivência. Em situações de dificuldades financeiras, a educação não é prioridade de quitação de dívidas para o cidadão brasileiro. Ainda segundo os autores, as prioridades são pagamentos relativos a financiamento de imóveis, seguido de financiamento de automóveis, despesas bancárias, cartão de crédito e comércio, saúde e, por fim, os gastos com educação. Como o aluno não sofre nenhuma restrição imediata por estar inadimplente, a educação não é considerada como uma dívida urgente.

Com a democratização do ensino superior no país, o aumento do número de IES e a forte concorrência no setor, percebe-se uma mudança no relacionamento entre as IES e alunos nos últimos anos. Os alunos passaram a demandar um serviço de qualidade cada vez maior, estrutura física moderna, atendimento satisfatório e atencioso, boa localização, e, principalmente, um preço que se encaixe em seu orçamento mensal (FREIRE E FREIRE, 2014). Os alunos deixaram de ser considerados como um produto das políticas e diretrizes institucionais, e passaram a se relacionar com as IES como clientes e fornecedores de serviço, respectivamente (COLOMBO, 2005). Assim, como toda empresa que mantém relação com um cliente, as IES correm o risco da inadimplência.

No setor de educação superior, inadimplência é um problema para a gestão financeira das empresas à medida que diminui as receitas das IES. Conseqüentemente, com menos caixa, as obrigações de curto prazo não são cumpridas e as reservas financeiras são consumidas, prejudicando a capacidade de investimento e competitividade organizacional (BEM et al., 2008).

A inadimplência faz com que muitas IES tenham que lançar mão de empréstimos bancários, encarecendo seu custo operacional, que acaba sendo repassado aos alunos na forma de aumento de mensalidade, impactando alunos adimplentes e inadimplentes e consolidando um ciclo vicioso, dado que em determinado momento, os alunos adimplentes não conseguirão honrar com suas obrigações devido ao preço elevado das mensalidades, tornando-se, assim, inadimplentes (HERLING et al., 2013).

Machado (2009), Herling et al. (2013) e Lemos, Ribeiro e Siqueira (2017) defendem que os principais obstáculos na gestão da inadimplência em IES estão relacionados à legislação educacional vigente, que protege o aluno, e a falta de adaptação das instituições à nova realidade do mercado, no que diz respeito às políticas de cobrança de débitos vencidos. De acordo com Freire e Freire (2014), a maior parte das IES esperam o final do semestre ou ano letivo, época de renovação de matrículas, para que o aluno ou seu responsável se apresente para a negociação, o que pode ser até mesmo um incentivo para que os alunos deixem de pagar mais mensalidades e consigam uma boa negociação ao final do semestre.

2.5.1. Legislação educacional vigente

De acordo com Freire e Freire (2014), devido à concepção constitucional da educação ser diretamente relacionada ao aspecto social, o Estado avalia que a concessão das instituições de ensino ao mercado não pode ser irrestrita, e, dessa forma, as IES jamais serão avaliadas como uma empresa pura que presta serviços em busca do lucro. Terão sempre uma legislação rigorosa, com objetivo de proteger o estudante e privilegiá-lo em sua busca pela educação superior.

A Lei nº 9.870/99 é extremamente relevante para a legislação educacional. Ela delimita alguns importantes detalhes no que diz respeito a cobrança de débitos vencidos entre contratante e contratado:

Art.5. Os alunos já matriculados, salvo quando inadimplentes, terão direito à renovação das matrículas, observado o calendário escolar da instituição, o regimento da escola ou cláusula contratual.

Art. 6. São proibidas a suspensão de provas escolares, a retenção de documentos escolares ou a aplicação de quaisquer outras penalidades pedagógicas por motivo de inadimplemento, sujeitando-se o contratante, no que couber, às sanções legais e administrativas, compatíveis com o Código de Defesa do Consumidor, e com os

artgs. 177 e 1.092 do Código Civil Brasileiro, caso a inadimplência perdure por mais de 90 dias.

O artigo 6º estabelece que nenhuma sanção pedagógica pode ser aplicada ao aluno inadimplente, incluindo a retenção de seus documentos escolares, como os de transferência para outra IES e histórico escolar, além de não ser permitida a proibição para realizar provas e frequentar aulas. Entretanto, o artigo 5º diz que o aluno já matriculado tem direito à rematrícula com exceção dos inadimplentes.

A Medida Provisória nº 2.173-24, de 23 de agosto de 2001 veio acrescentar mais um parágrafo no artigo 6º da Lei nº 9.870/99, que estabelece:

Art. 2º - O art. 6º da Lei 9.870, de 1999, passa a vigorar acrescido do seguinte § 1º, renumerando-se os atuais §§ 1º, 2º e 3º para §§ 2º, 3º e 4º:

§ 1º - O desligamento do aluno por inadimplência somente poderá ocorrer ao final do ano letivo ou, no ensino superior, ao final do semestre letivo quando a instituição adotar o regime didático semestral. (CÓDIGO DE DEFESA DO CONSUMIDOR, 2004. p.356)

De acordo com Andrade et al. (2008), essa Medida Provisória veio preencher uma lacuna da Lei 9.870/99, que não explicou claramente se a contratada poderia recusar a rematrícula do inadimplente, se ele ainda estivesse em débito com a instituição de ensino. Assim, ficou ratificado que, de fato, só é possível realizar a renovação de matrícula para o semestre letivo subsequente se o aluno estiver adimplente com as mensalidades de seu curso.

2.5.2. Ações contra a inadimplência – políticas de cobrança

A prática mais adotada pelas IES para combater a inadimplência é a negociação das dívidas através de políticas de cobrança, que buscam a recuperação de créditos (ANDRADE et al., 2008). Diferente de outros setores de negócio, as IES não estão legalmente aptas a deixarem de prestar o serviço no momento que o cliente/aluno se torna inadimplente, o que faz com que a política de cobrança seja complexa, uma vez que o aluno inadimplente prioriza dívidas mais urgentes em detrimento das educacionais.

De acordo com Andrade et al. (2008), as políticas de cobrança em IES devem ser resguardadas por um contrato de prestação de serviços educacionais assinado pelo aluno no ato de sua matrícula, e que contenha uma cláusula de inadimplência

bem rígida, deixando claro ao contratante quais as consequências decorrentes do não pagamento da mensalidade. Além disso, a política de cobrança institucional deve oferecer uma boa forma de negociação, que atraia os alunos a quitarem seus débitos.

De acordo com Andrade et al. (2008) e Herling et al. (2013), a política de cobrança é composta por quatro fases, a primeira delas é a matrícula, em que ocorre a assinatura do contrato de prestação de serviços educacionais entre a IES e o aluno e que deve ser bem elaborado, uma vez que é a principal prova para realização da cobrança judicial ou extrajudicial. Neste momento, o ideal seria realizar uma análise financeira dos contratantes, uma vez que a IES irá conceder um crédito para o aluno, e esta análise ajudaria a instituição a reduzir o risco de inadimplência. Entretanto, a maior parte das IES, pressionadas pelas questões comerciais de aumentar o número de matrículas, acabam negligenciando esta etapa.

A segunda fase é a da cobrança interna, onde uma equipe da própria IES fica encarregada de fazer a cobrança dos inadimplentes, de acordo com as regras de negócio. Nesta etapa, lança-se mão de todas as alternativas para que o aluno quite sua dívida e renove sua matrícula, evitando que, possivelmente, venha a evadir. A terceira fase é a da cobrança externa, feita por empresas especializadas no setor. Geralmente, essas empresas ficam encarregadas da cobrança de débitos mais antigos, de difícil recuperação, que demandaria mais tempo das equipes internas. Por fim, a última fase é a da cobrança jurídica, realizada por escritórios de advocacia que conheçam a legislação educacional, por tratar-se de um processo cauteloso que demanda conhecimento técnico para mitigar riscos de futuras ações judiciais contra as IES (HERLING et al., 2013).

Cabe ainda ressaltar que muitas IES lançam mão do uso de sistemas de consulta ao crédito, como o SPC e Serasa Experian para realizar a negativação daqueles alunos que estejam em débito com a IES, de acordo com as cláusulas previstas no contrato de prestação de serviços educacionais para tal inclusão. Trata-se de uma forma de estimular que os alunos inadimplentes quitem suas dívidas com a instituição, uma vez que a negativação restringe o acesso do devedor ao crédito, como financiamentos, compras a prazo, empréstimos, entre outras limitações (ANDRADE et al., 2008).

2.5.3. Fatores que impactam a inadimplência

A inadimplência causa transtornos e prejuízos às IES e, por isso, é extremamente importante identificar os fatores que a origina, sendo possível assim traçar estratégias preventivas (HERLING et al., 2013). De acordo com Andrade et al. (2008), os principais fatores externos que impactam a inadimplência nas IES são a situação econômica do país, a legislação educacional vigente, a má concessão de crédito por parte das IES e a falta de adaptação das IES à nova realidade do mercado, que exige uma resposta cada vez mais rápida, principalmente no que diz respeito a cobrança de créditos a receber.

Santos (2003) atribui à situação econômica do país a principal causa para a inadimplência nas IES. O nível de desemprego e endividamento da população e a volatilidade das taxas de juros, por exemplo, afetam diretamente a capacidade pagadora dos tomadores de crédito. Para Andrade et al. (2008), o nível de desemprego é uma questão social relacionada à política governamental do país e, neste caso, o devedor é considerado uma vítima. Com relação ao endividamento da população, segundo os mesmos autores, trata-se de uma consequência de uma economia estável, da busca do brasileiro por uma qualidade de vida melhor e da falta de conhecimento sobre planejamento financeiro. O endividamento pode ser considerado como um retorno das famílias ao consumo, ou pelo menos, um sinal de aumento de confiança para com a situação econômica do país. A redução dos juros também facilita o consumo das famílias, uma vez que é possível obter crédito mais barato e com mais facilidade.

Com relação a má concessão de crédito, as IES, em sua maioria, não realizam qualquer tipo de análise de crédito prévia do aluno para prever seu potencial risco de se tornar inadimplente e poder realizar ações preventivas. Algumas IES lançam mão apenas de consultas a sites como SPC e Serasa Experian para conhecer o histórico financeiro de crédito do aluno, mas sem a intenção de restringir o acesso aos serviços. Com isso, as IES se sujeitam a um maior risco de inadimplência, uma vez que não conhecem os potenciais maus pagadores.

Outro aspecto relevante do atual cenário educacional no país que impacta os índices de inadimplência nas IES é a crescente redução do número de vagas ofertadas pelo FIES. Com isso, as IES passaram a ser as principais responsáveis por financiar seus alunos através de programas próprios, como medida para evitar o

abandono por dificuldades financeiras, incidindo sobre elas todo o risco de inadimplência atrelado a programas de concessão de crédito (SEMESP, 2019). Além disso, a legislação educacional vigente impede que as IES deixem de prestar o serviço no momento que o aluno se torna inadimplente, somente ao final de cada semestre, o que pode levar um aluno devedor a estudar quase seis meses sem pagar as mensalidades. Dada essas circunstâncias, o acompanhamento e controle da inadimplência torna-se crucial no contexto atual para as IES que desejam manter uma gestão financeira eficiente, em um mercado cada vez mais dinâmico e competitivo.

Lemos, Ribeiro e Siqueira (2017) propõem que a inadimplência nas IES seja analisada sob uma perspectiva mais ampla, que extrapole exclusivamente os fatores externos e financeiros como determinantes da capacidade pagadora dos alunos. Os autores procuram explicar a inadimplência através de fatores internos dos próprios alunos, como por exemplo, análise de seus perfis, suas características socioeconômicas e situação acadêmica.

Os autores supracitados apresentam sete fatores como sendo influentes na capacidade pagadora dos alunos. No que diz respeito aos aspectos acadêmicos, o primeiro deles é a sua origem acadêmica, relacionada com o histórico acadêmico do aluno e ao fato dele ter realizado o ensino médio em uma escola pública ou particular, o que pode vir a influenciar em sua capacidade pagadora. Em seguida, o rendimento acadêmico recente pode ser uma variável influente no risco de um aluno se tornar inadimplente, uma vez que o fato do aluno apresentar baixo rendimento acadêmico poderá contribuir para o fortalecimento da falta de interesse em efetuar os pagamentos das mensalidades nas datas aprazadas, dando prioridade aos outros pagamentos que despertem maior interesse. A frequência em sala de aula (assiduidade) também é um fator acadêmico relacionado a motivação do aluno para com a IES e considerado pelos autores como relevante no estudo da inadimplência. A constância ou inconstância às salas de aulas poderá gerar o sentimento de pertencer ou não pertencer ao ambiente acadêmico, além de deixar o aluno propenso a desviar a destinação dos recursos financeiros por não estar absorvendo os conhecimentos de forma proporcional aos valores despendidos mensalmente (LEMOS; RIBEIRO; SIQUEIRA, 2017).

Com relação aos fatores demográficos que impactam a inadimplência, os autores mencionam o gênero, uma vez que estudos realizados pelo Serasa Experian

(2019) apontam diferenças a respeito do perfil de endividamento entre homens e mulheres. Segundo a fonte supracitada, os homens são mais devedores no segmento de bancos e cartões de crédito, e as mulheres, têm mais dívidas atrasadas em contas de água, luz e gás. Outro fator demográfico que influencia a inadimplência é o local de residência, dado que a renda per capita varia de acordo com o local de moradia. A última variável demográfica apresentada pelos autores é o estado civil. Trata-se de uma variável relacionada ao fato do aluno ser solteiro, casado ou divorciado, e que pode influenciar a sua capacidade pagadora à medida que as necessidades de dispêndios com outras obrigações financeiras podem ser diferentes de acordo com o seu estado civil.

Por fim, a ocupação profissional também é considerada um fator preditor da inadimplência, à medida que, como já apresentado previamente neste estudo, o desemprego é considerado como uma das suas primeiras causas. Além disso, a natureza da ocupação profissional, seja um emprego autônomo, ou fixo em empresa privada ou pública, pode vir a influenciar a capacidade pagadora do aluno, dado a imprevisibilidade dos mercados.

2.6. Retenção no setor de educação superior

Shajahan (2006) define retenção como a habilidade de uma companhia em reter seus consumidores já adquiridos e oferecer para um cliente, além do produto para compra, um padrão de relacionamento dentro de um período específico. Ahmad e Buttle (2001) colocam que a retenção de clientes tem sido vista como uma imagem espelhada da deserção de clientes da empresa, ou seja, altos índices de retenção significam baixas taxas de saída da empresa.

Para Henning-Thurau e Hansen (2000), a retenção, em grande parte das vezes, é motivada por ímpeto emocional. Desta forma, conhecer desejos e necessidades do cliente, bem como as razões implícitas para manter ou sair de um relacionamento, torna-se imprescindível.

Para Astin et al. (2012), retenção, no contexto educacional superior, é a habilidade de uma IES de graduar com sucesso os estudantes que se matriculam inicialmente na instituição. De acordo com Oliveira e Barbosa (2016), a retenção é uma problemática grave que limita a efetiva expansão da educação superior. Ela exige repasses financeiros além do esperado, já que a instituição é levada a manter

toda uma estrutura física, tecnológica e de recursos humanos para atender um conjunto de discentes que permanecerá além do tempo regular.

Bergamo et al. (2011) colocam a questão da evasão de alunos dentro da educação como uma “epidemia”, que vem afetando a sobrevivência de instituições educacionais em todos os seus níveis. Para Tinto (1975), as principais causas da evasão estudantil estão relacionadas às características individuais daqueles que evadem seus cursos, seguido das características relacionadas ao relacionamento entre esses indivíduos e as IES e, por fim, as características das próprias IES, que podem estar relacionadas à evasão. No âmbito individual, além das características pessoais do próprio aluno, o autor cita também a sua estrutura familiar, suas experiências acadêmicas anteriores e suas expectativas em relação a seu futuro profissional como questões que influenciam o grau de compromisso dos alunos em concluir cursos superiores.

Tinto (1999) afirma que os índices de evasão são efetivamente maiores em alunos recém-matriculados ou nos primeiros anos, caindo a cada ano que o estudante renova a sua matrícula até a conclusão do curso superior. O autor afirma, ainda, que o tratamento de retenção visando o estudante de primeiro ano é essencial para a persistência do mesmo, sugerindo que as instituições implantem seminários para alunos ingressantes, acompanhamento acadêmico, assim como programas de imersão na IES, para que os alunos se sintam motivados e integrados ao ambiente institucional.

De acordo com Bergamo et al. (2012), para gerenciar a retenção dos estudantes, as IES devem estabelecer políticas de retenção, cujo foco é avaliar a situação da retenção institucional, determinar suas causas e criar ações para reduzir os índices, estabelecendo um ambiente orientado aos alunos, prestando aconselhamento e orientação. Os alunos devem se sentir acolhidos e integrados ao ambiente acadêmico e social da IES.

Para Costa e Gouveia (2018), algumas discussões pressupõem a existência de uma variedade de tipos de retenção de estudantes que permanecem na instituição ou não. Em função desta variedade, deve-se ter uma atenção em especial na distinção das mesmas, visto que cada uma delas pode estar associada a diferentes fatores e causas. Ahmad e Buttle (2001) identificam três tipos de retenção: Retenção dentro de um curso, retenção institucional e retenção do sistema. Na primeira, os esforços da IES direcionam-se a manter o aluno em seu curso. Na retenção

institucional, as IES procuram evitar que o aluno se transfira para outra instituição, e na retenção do sistema, os esforços são direcionados a evitar que o aluno abandone o ensino superior. Do ponto de vista das IES, as duas últimas representam uma perda já que, na mudança de curso, o aluno continua na instituição, apenas consumindo um serviço diferente.

2.6.1. Fatores-chaves relacionados à retenção de estudantes

Tinto (1975) ressalta que dentre as características mais utilizadas na segmentação de estudantes dentro da educação superior estão os aspectos demográficos, que envolvem a raça e a etnia, gênero, estado civil, idade e classe social, e os aspectos educacionais, que envolvem ajuda financeira (estudante bolsista ou não), performance acadêmica, forma de matrícula (vestibular, transferência interna ou externa), escola de procedência (escola pública, particular ou comunitária), turno (diurno ou noturno), integralidade. O estudo destes aspectos é de suma importância para que as IES consigam entender o que levam os estudantes a evadirem seus cursos e, assim, promover ações de retenção focadas no peso destes fatores.

Baseado nos inúmeros modelos existentes na literatura que procuram entender o fenômeno do abandono estudantil, Costa e Gouveia (2018) apresentaram fatores-chave que influenciam, direta ou indiretamente, a permanência dos alunos em seus cursos:

- **Preparação acadêmica:** A qualidade da instrução anterior de um aluno pode influenciar significativamente se um aluno terá ou não sucesso em uma IES.
- **Integração social:** Os alunos que se envolvem em atividades dentro e fora do ambiente acadêmico estão menos propensos à abandonarem seus cursos.
- **Integração acadêmica:** O envolvimento do aluno no ambiente acadêmico, medido através da relação com docentes, participação em grupos de estudo e horas dedicadas aos estudos são fatores que demonstram a motivação em concluir o curso.

- Compromisso com a IES: Percepção que o estudante tem da instituição, através do quão engajado ele está no ambiente acadêmico e social da IES.
- Compromisso com o objetivo: Percepção que o estudante tem da qualidade do curso, e utilidade do diploma frente ao esforço necessário para manter seu vínculo.
- Ambiente: O ambiente inclui fatores externos como finanças, trabalho e família, que podem afetar o aluno e o seu desempenho durante a jornada acadêmica.
- Características demográficas dos alunos: Idade, profissão, gênero, estado civil, naturalidade e etnia são fatores que podem afetar a permanência dos alunos nas IES.

2.6.2. Lealdade no setor de educação superior

Entre os estudos que visam entender a lealdade entre os alunos e as IES, um dos mais citados é o de Hennig-Thurau et al. (2001), que desenvolveram o modelo RQSL (*Relationship Quality-based Student Loyalty*), baseado no modelo de retenção, de Tinto (1975), e no de Qualidade do Relacionamento, de Hennig-Thurau e Klee (1997). No estudo, considera-se que a lealdade do estudante é importante para a IES por reter alunos e desenvolver sólida base financeira, possibilitando vantagens competitivas, uma vez que estudantes leais podem influenciar positivamente a qualidade do ensino e do serviço da IES, pela participação ativa e comportamento comprometido.

No modelo RQSL, a lealdade é função três constructos: a percepção dos estudantes da qualidade da educação (a qualidade do serviço), a confiança do estudante nos atores da instituição (funcionários acadêmicos ou administrativos) e o comprometimento do aluno com a instituição (comprometimento emocional, comprometimento com objetivos do aluno, integração acadêmica e social, inserção no mercado de trabalho, presença ativa dos familiares e atividades não-universitárias no campus) (PEREIRA, 2017).

Além disso, conforme indicam os estudos apresentados por Kumar (2018), percebe-se que a lealdade do aluno com a IES permanece mesmo após a conclusão do curso, através da criação de uma relação de longo prazo que transcende as trocas

transacionais. Alunos leais fazem propaganda boca-a-boca positiva, indicações e estão mais propensos a ingressarem em outros cursos de graduação ou cursos de extensão e pós-graduação, evidenciando que a criação da lealdade entre aluno e IES trata-se de um aspecto de suma importância, tanto no que diz respeito a retenção, como também na captação de novos alunos e/ou alunos egressos para novos cursos.

2.7. Mineração de dados e *Machine Learning*

Mineração de dados é um termo frequentemente usado para definir o processo de extração de informações úteis de um vasto banco de dados de informações dos clientes (PEPPERS E ROGERS, 2011). Com a evolução do relacionamento e interação entre clientes e empresas, a abordagem de mineração de dados também sofreu algumas mudanças. Na fase chamada pelos autores supracitados como pré-interativa, as técnicas de mineração de dados eram utilizadas para descobrir informações sobre para qual tipo de clientes determinadas ofertadas deveriam ser feitas, respondendo à pergunta: “Quem é o cliente com mais chances de comprar determinado produto?”. Atualmente, a pergunta que as empresas engajadas na gestão de relacionamentos contínuos e interativos com clientes devem responder é “Qual é o próximo produto que determinado cliente vai querer comprar?”. Percebe-se uma mudança de foco no produto para um foco maior no conhecimento de cada cliente individualmente, suas necessidades e comportamentos. O principal desafio para as empresas que trabalham com uma grande quantidade de dados de seus clientes é aproveitar essa oportunidade dada pela disponibilidade de dados e transformá-los em conhecimento útil para a empresa.

Em geral, a mineração de dados envolve ao menos uma das seguintes atividades: classificação, estimação e predição – atividades de mineração direta; agrupamento por afinidades, *clustering*, descrição e visualização – mineração indireta (HAN E KAMBER, 2006). Junto com essas atividades, diferentes algoritmos podem ser utilizados, como regressão logística, árvores de decisão e redes neurais, que serão utilizados neste trabalho.

Estas mesmas técnicas são aplicadas no domínio do que é conhecido como *machine learning* e muitas vezes essas terminologias se confundem. No entanto, é possível diferenciá-los pelos objetivos da análise realizada. Enquanto que em

mineração de dados o objetivo é identificar padrões ainda não conhecidos entre as variáveis utilizadas, em *machine learning*, o objetivo é aprender continuamente com novos dados a fim de melhorar a previsão sobre uma determinada variável. Assim, em *machine learning* o pesquisador irá incluir, dentre os atributos para gerar o modelo, variáveis que já são suspeitas de causar impacto na variável-alvo e o objetivo é, dado esses atributos, avaliar qual o *output* que o modelo retorna para cada instância da base de dados. Em um trabalho de mineração de dados, atributos diversos sobre os quais suas relações são desconhecidas são avaliados, em busca de relações ocultas, previamente desconhecidas (PEREIRA, 2017). Portanto, considerando o objetivo deste trabalho, pode-se dizer que se trata de uma aplicação de *machine learning*, onde os algoritmos serão utilizados como forma de se coletar as informações relevantes dos alunos da IES avaliada e identificar aquelas que influenciam a capacidade pagadora dos mesmos.

No cenário do ensino superior brasileiro, as técnicas de *machine learning* encontram um vasto campo de aplicação, ainda pouco explorado. Estudos de Vianna et al. (2017) sobre o uso do *machine learning* na predição do *ranking* de universidades, Pereira (2017) sobre o uso de técnicas de *machine learning* para a evasão de alunos, Cechinel, Araujo e Detoni (2015) sobre a predição de reprovação de alunos em cursos à distância e Gottardo, Kaestner e Noronha (2012) sobre o desempenho de alunos de cursos a distância, são alguns exemplos na literatura brasileira do uso de técnicas de *machine learning* na educação superior. A inadimplência, especificamente, pouco foi estudada neste setor utilizando estes modelos, sendo um campo promissor que apresenta muitas possibilidades de investigação a serem exploradas.

3 Metodologia

Este estudo aborda a inadimplência no ensino superior privado como um problema de gestão financeira e de marketing de relacionamento. Para tanto, é necessário avaliar informações específicas de clientes, permitindo que suas necessidades sejam entendidas e que eles sejam tratados de forma individualizada, o que aumenta a eficácia das ações de marketing, particularmente ações de retenção. Optou-se neste estudo pela utilização de uma abordagem quantitativa, usufruindo-se do acesso aos conjuntos de dados de sistemas financeiro e acadêmico de uma IES, disponibilizados com exclusividade por sua equipe de gestores especificamente para este trabalho. Bases de dados internas de clientes de empresas podem proporcionar ricas oportunidades de análise de comportamento do consumidor, porém, em geral, a academia possui escasso acesso a esse tipo de informação para a atividade de pesquisa.

A seguir, é apresentada uma descrição sobre a IES avaliada, o tipo de pesquisa realizado, os dados coletados, e os procedimentos utilizados para tratar e analisar os dados.

3.1 A Instituição de Ensino Superior avaliada

A instituição avaliada está há 30 anos no mercado de educação superior privada do Rio de Janeiro e possui mais de 13 mil alunos, distribuídos entre os *campi* do Rio Comprido e Méier. O perfil do aluno da IES é um público de baixa renda de grandes centros urbanos. Atualmente, são oferecidos 17 cursos de graduação presencial, 10 cursos de pós-graduação, além de cursos de graduação à distância e mestrado profissional.

O índice de inadimplência da IES avaliada é de 20% da base de alunado com débitos de uma ou mais mensalidades. Pretende-se, como resultado deste estudo, que a IES possa aplicar os modelos de *machine learning* como uma ferramenta de apoio gerencial às tomadas de decisão acerca das políticas de controle de

inadimplência, obtendo resultados mais efetivos na redução dos índices institucionais.

O acesso aos dados dos alunos da IES se deu através de uma solicitação formal à diretoria executiva, que uma vez aprovada pelo setor jurídico da empresa, requereu apenas que os dados fossem descaracterizados, para garantir a privacidade de informação dos alunos.

3.2 Tipo de Pesquisa

Este estudo utilizou uma abordagem quantitativa de análise, aplicando modelos de *machine learning* para classificação de uma base de dados secundários, a fim de encontrar modelos que expliquem a inadimplência na IES avaliada. A base de dados possui informações referentes a uma amostra representativa de aproximadamente 5 mil alunos da instituição avaliada que estavam ativos nos anos de 2017 e 2018. É importante ressaltar que a escolha de uma base de dados somente com alunos ativos nos anos de 2017 e 2018 (4 semestres) se explica pela relevância de poder realizar uma análise comparativa e evolutiva do mesmo aluno ao longo destes 4 semestres. Foram criadas variáveis adicionais que englobassem os resultados dessa análise. A escolha dos anos de 2017 e 2018 se justifica pelo fato que, no momento da extração dos dados ainda não era possível obter todas as informações do ano de 2019.

Para os testes e simulações realizadas, foi utilizada a ferramenta de análise de dados WEKA, e aplicou-se três tipos de algoritmos de classificação, a regressão logística, árvores de decisão e redes neurais. De forma genérica, o que esses algoritmos fazem é avaliar os atributos de cada instância e sua respectiva classe para classificá-los de acordo com o seu risco de se tornar inadimplente (PEREIRA, 2017). Uma consequência importante dessa metodologia é que os resultados dos modelos não são capazes de dizer em que momento o aluno irá se tornar inadimplente. Os processos de classificação apenas indicarão a probabilidade de inadimplência do aluno em algum momento, dadas as informações financeiras, acadêmicas e sociais disponibilizadas.

3.3 Definição da variável-alvo

Como o objetivo deste estudo é gerar um modelo preditivo de classificação do risco de inadimplência para alunos ativos, é importante definir o conceito de inadimplência para a IES avaliada. **Inadimplente** é aquele aluno que possui de um a seis débitos de mensalidade. A partir do quinto dia após a data de vencimento do boleto de cobrança, o aluno é considerado inadimplente. O limite da inadimplência são seis débitos, uma vez que de acordo com a legislação educacional vigente, as IES não são obrigadas a renovar a matrícula para o semestre subsequente dos alunos devedores. **Adimplente** é aquele aluno que não possui nenhum débito de mensalidade.

A variável-alvo foi criada a partir da variável “quantidade de débitos em 2018.2”. As variáveis relacionadas a quantidades de débitos nos outros três semestres estudados (2018.1, 2017.2 e 2017.1) farão parte da modelagem como atributos a serem testados.

Outra observação relevante acerca da modelagem é que, como mencionado anteriormente neste estudo, diferente de outros serviços, os educacionais não podem deixar de serem prestados por motivos de inadimplência, até o final do semestre vigente. Dessa forma, a situação de inadimplência dos alunos apresentada neste trabalho será referente ao final de cada semestre. Além do aspecto da legislação educacional, como a regra de inadimplência na IES avaliada é cinco dias após o vencimento do boleto, a situação do aluno pode variar ao longo de um semestre, sendo mais coerente utilizar a situação de inadimplência ao final do semestre, que é o prazo limite para que o aluno esteja adimplente e apto para renovar a matrícula.

Com relação às proporções entre as classes da amostra, em geral, o objetivo que se deseja atingir é que a classe rara, ou seja, aquela com menor frequência na amostra, corresponda a algo entre 10% e 40% da base para a modelagem. Frequências entre 20% e 30% em geral dão bons resultados (BERRY E LINOFF, 2000). Quando a amostra não apresenta uma proporção adequada entre classes, é necessário realizar o procedimento de *oversampling*, cujo objetivo é ajustar as proporções entre classes, através da criação de uma nova base de dados. Neste estudo, a proporção entre as classes na amostra utilizada está 38% dos alunos como

adimplentes e 62% inadimplentes, não sendo necessário utilizar o procedimento de *oversampling*.

3.4 Dados coletados

Na tabela 3.1 estão dispostas as variáveis primárias coletadas nos bancos de dados da IES.

Tabela 3.1 - Tabela de Dados Coletados

Categoria	Informação	Comentário
Sociodemográficas	Matrícula	Todas as informações dos bancos de dados acadêmicos estão associadas pelo número de matrícula, por isso este foi utilizado como código identificador de cada aluno neste estudo.
	Gênero	Masculino ou feminino
	Estado civil	{ 1,2,3,4 } onde 1-casado, 2-divorciado, 3-solteiro e 4-viúvo
	Idade	Calculado no momento da extração dos dados em setembro/2019
	Possui filhos?	Sim ou não
	Quantidade de pessoas que residem com aluno	De 0 a 5 pessoas
	Participação vida econômica	Variáveis binárias que buscam entender se o aluno é responsável financeiramente
	Atividade remunerada	Variáveis binárias que buscam entender a natureza da atividade remunerada exercida pelo aluno: autônomo, estágio, emprego fixo particular ou estatal
	Contribuição renda familiar	Próprio aluno, pai, mãe, cônjuge ou outra pessoa
	Renda familiar	Até 02 salários mínimos, de 02 a 04 salários mínimos ou maior do que 04 salários mínimos
Acadêmicas	Ensino médio	Colégio público ou particular sem bolsa, com bolsa parcial ou integral
	Disciplinas	Quantidade de disciplinas que o aluno se matriculou no semestre
	Faltas	Quantidade total de faltas do aluno por semestre
	Reprovações	Quantidade total de reprovações por semestre

	Reprovações por falta	Quantidade de reprovações por falta a cada semestre
	Reprovações por nota	Quantidade de reprovações por nota a cada semestre
	Aprovações	Quantidade de aprovações por semestre
	Trancamentos	Quantidade de trancamentos de disciplinas por semestre
Financeiras	Valor pago	Valor da semestralidade paga
	Quantidade de débitos	Quantidade de débitos (de 0 a 6) nos 4 semestres analisados

3.5 Preparação dos dados

A seguir serão apresentadas algumas técnicas e práticas utilizadas na preparação de uma base de dados para a extração do conhecimento por meio da utilização de modelos de *machine learning*. Tais procedimentos visam garantir a consistência dos dados e uma representação dos mesmos que facilite a compreensão pelos modelos que serão aplicados. Outros procedimentos podem ser encontrados na literatura (BALLOU E TAYI, 1999; BERRY E LINOFF, 2000; PYLE, 1999) para serem aplicados de acordo com os dados disponíveis e em função da natureza da questão-problema a ser solucionada. Os métodos abordados nesta seção serão utilizados na modelagem proposta para melhoria da efetividade dos modelos de *machine learning* em prever o risco de um aluno se tornar inadimplente.

3.5.1. Tratamento e limpeza

Os dados reais tendem a ser incompletos, fora de padrões e inconsistentes. A limpeza dos dados visa detectar e remover anomalias presentes na base de dados com o objetivo de melhorar a sua qualidade. Esta rotina implica em esforços de preencher os valores ausentes (*missing values*) e padronizar dados (*standardization*), enquanto identifica valores fora do padrão (*outliers*) e corrige inconsistências nos dados (CÔRTEZ; PORCARO; LIFSCHITZ, 2002).

Os valores ausentes (*missing values*) são caracterizados por campos que não possuem nenhum valor armazenado, e que podem ser importantes para o modelo. Os métodos utilizados nestes casos são: ignorar o registro, indicado quando os dados são abundantes, e especialmente pobre quando o percentual de valores

ausentes varia consideravelmente, preencher o valor manualmente, o que demanda muito tempo e torna-se impraticável em base de dados muito grandes, usar uma constante global para todos os dados ausentes, utilizar um atributo médio para preencher os valores ausentes, utilizar um atributo médio pertencente a mesma classe do valor ausente, e, por fim, utilizar o valor de maior probabilidade para preencher os valores ausentes. Neste trabalho, o foco da coleta de dados foi obter uma amostra que tivesse todos os valores preenchidos, uma vez que é de grande relevância poder comparar alunos que estivessem na mesma situação nos semestres analisados, até mesmo para que variáveis derivadas pudessem ser criadas.

Valores fora de padrão (*outliers*) são um erro aleatório ou variação acentuada na medição de uma variável (CÔRTEZ; PORCARO; LIFSCHITZ, 2002). Alunos com idade acima de 60 ou abaixo de 18 anos e valores pagos de semestralidade fora do intervalo de dois desvios padrões da média foram identificados como *outliers* e desconsiderados neste estudo. Este processo busca evitar que os valores incorretos atrapalhem a compreensão dos dados pelos modelos, levando-os a conclusões errôneas.

3.5.1.1. Variáveis *dummy*

A presença de variáveis discriminantes categóricas também exige um processo de transformação. As variáveis binárias, também conhecidas como *dummy*, são definidas como aquelas que assumem somente dois valores 0 (negativo) e 1 (positivo) (GUJARATI, 1970). São utilizadas como forma de acomodar as variáveis não-métricas. Neste estudo, utilizou-se variáveis binárias para classificação de variáveis sociodemográficas dos alunos, conforme apresenta a tabela 3.2.

Tabela 3.2 - Tabela de Variáveis Binárias

Informação	Variável binária
Participação_vida_econômica_=_você_não_trabalha_e_seus_gastos_são_custeados	{0,1}
Participação_vida_econômica_=_você_trabalha_e_é_responsável	{0,1}

Participação_vida_econômica_=_você_trabalha_mas_nao_é_independent e	{0,1}
Participação_vida_econômica_=_você_trabalha_e_é_independente_finan ceiramente	{0,1}
Atividade_remunerada_=_emprego_autônomo	{0,1}
Atividade_remunerada_=_sem_atividade	{0,1}
Atividade_remunerada_=_estágio	{0,1}
Atividade_remunerada_=_emprego_fixo_particular	{0,1}
Atividade_remunerada_=_emprego_fixo_federal_estadual_municipal	{0,1}
Atividade_remunerada_=_nenhuma_das_opções_anteriores	{0,1}
Contribuição_renda_familiar_=_pai	{0,1}
Contribuição_renda_familiar_=_mãe	{0,1}
Contribuição_renda_familiar_=_cônjuge_ou_companheiro	{0,1}
Contribuição_renda_familiar_=_você_mesmo	{0,1}
Contribuição_renda_familiar_=_outra_pessoa	{0,1}
Filhos_=_não	{0,1}
Renda_familiar_=_ate_02_salários_mínimos	{0,1}
Renda_familiar_=_02_a_04_salários_mínimos	{0,1}
Renda_familiar_=_acima_de_04_salários_mínimos	{0,1}

Ensino_médio=_escola_particular_com_bolsa_integral	{0,1}
Ensino_médio=_escola_particular_sem_bolsa	{0,1}
Ensino_médio=_escola_particular_com_bolsa_parcial	{0,1}
Ensino_médio=_escola_pública	{0,1}

3.5.2. Adição de variáveis derivadas

Na preparação dos dados, pode ser interessante a criação de novas variáveis a partir das variáveis existentes na base de dados. A criação de novas variáveis tem como objetivo enfatizar certos aspectos dos clientes que podem ajudar e enriquecer a modelagem. Na tabela 3.3 são apresentadas as variáveis criadas a partir dos dados primários disponibilizados pela IES.

Tabela 3.3 - Tabela de Variáveis Derivadas

Categoria	Informação	Comentário
Financeiras	Variação valor pago	Variação do valor pago ao longo dos 4 semestres com o objetivo de perceber tendência positiva ou negativa e relacionar com a inadimplência
	Média valor pago	Valor médio pago ao longo dos 4 semestres analisados
Acadêmicas	Variação quantidade disciplinas	Variação da quantidade de disciplinas cursadas a cada semestre com o objetivo de perceber tendência positiva ou negativa e relacionar com a inadimplência
	Média quantidade de disciplinas	Quantidade média de disciplinas cursadas nos 4 semestres analisados
	Média de falta por disciplina	Quantidade média de faltas em relação ao número de disciplinas cursadas

Média de reprovações	Quantidade média de reprovações nos 4 semestres analisados
Percentual de reprovações	Percentual médio de reprovações em relação ao número de disciplinas cursadas
Média reprovações por falta	Quantidade média de reprovações por falta nos 4 semestres analisados
Média reprovações por nota	Quantidade média de reprovações por nota nos 4 semestres analisados
Percentual de trancamentos	Percentual médio de trancamentos em relação ao número de disciplinas cursadas
Média de trancamentos	Quantidade média de trancamentos nos 4 semestres analisados
Percentual aprovações	Percentual médio de aprovações em relação ao número de disciplinas cursadas
Média de aprovações	Quantidade média de aprovações nos 4 semestres analisados

Com relação às variáveis financeiras, foram criadas variáveis adicionais relacionadas a variação média do valor pago pelo aluno durante os 4 semestres analisados, além do valor médio pago neste período. O objetivo da criação destas variáveis foi a possibilidade de uma análise horizontal que reflita o histórico financeiro do aluno ao longo destes quatro semestres e sua relação com um perfil de inadimplência/adimplência, e não somente um retrato de um único semestre. Com essas variáveis, é possível, por exemplo, associar uma variação positiva ou negativa, isto é, um aumento ou redução do valor pago ao longo dos semestres analisados, com um possível risco de inadimplência, o que sinaliza para a gestão da IES uma maior atenção para variações no valor pago de semestralidade, que podem significar maior risco de inadimplência.

As variáveis acadêmicas criadas estão relacionadas às médias, percentuais e variações ao longo dos semestres analisados de variáveis acadêmicas obtidas primariamente no banco de dados da IES. O objetivo da criação destas variáveis foi a possibilidade de se analisar tendências que possam estar relacionadas com o risco de um aluno de tornar inadimplente. Por exemplo, se a variação da quantidade de disciplinas que o aluno se matriculou ao longo dos semestres analisados for

identificada como uma variável relevante para o risco de inadimplência, cabe aos gestores ficarem atentos tanto para variações positivas, como negativas da quantidade de disciplinas. Ou ainda, se o percentual de trancamentos for identificado como uma variável relevante para o modelo, os gestores devem promover ações pedagógicas a fim de reduzir a quantidade de trancamentos por semestre.

3.6 Análise dos dados

Este estudo utilizou o software de acesso gratuito WEKA para criar os modelos preditivos. Esse software possui um banco de algoritmos mais usados em processos de aprendizado de máquinas e permite criar, testar e comparar a qualidade modelos (FRANK et al., 2009). Utilizou-se três algoritmos de *machine learning* para classificação: regressão logística (WILSON E LORENTZ, 2015; HOSMER E LEMESHOW, 2010), árvore de decisão e redes neurais (HAYKIN, 2007), cujos detalhes dos resultados serão explorados no capítulo 4.

O tipo de problema que este estudo trabalha é o de classificar um aluno como adimplente ou inadimplente, de acordo com suas características acadêmicas, financeiras e sociais. Neste caso, utiliza-se uma abordagem conhecida como treinamento supervisionado, uma vez que os dados já estão classificados (0 – adimplente e 1 – inadimplente) e o algoritmo apenas aprende a relacionar os atributos àquela classe previamente definida. Em uma abordagem não supervisionada, não há classe definida e o algoritmo cria grupos a partir de associações “descobertas” – esse é o tipo de abordagem utilizado para construção de *clusters* por exemplo, ou de regras associativas (PEREIRA, 2017).

As acurácias *out-of-sample* dos modelos e suas capacidades de generalização foram testadas com o método de validação cruzada conhecido como *k-fold cross-validation* (WITTEN; FRANK; HALL, 2011), com $k = 10$. Essa técnica consiste na divisão aleatória do banco de treinamento em k partes de tamanhos iguais, em que $k-1$ irá compor os dados de treinamento para o ajuste de modelos, e a outra parte ficará reservada para a estimativa de sua performance. Esse procedimento é realizado k vezes e, ao fim, tem-se uma estatística da quantidade de previsões sobre a classe corretamente realizadas sobre o total de previsões. Não há uma regra precisa para a escolha de k , embora a divisão dos dados em 5 ou 10 partes seja mais comum.

4 Análise dos resultados

Este capítulo tem por objetivo apresentar a análise dos dados coletados na pesquisa quantitativa e a discussão dos resultados frente ao referencial teórico. Nele serão apresentados e comparados os modelos obtidos através do uso de algoritmos de *machine learning*.

4.1 Modelos gerados

Nas tabelas e gráficos a seguir os algoritmos estão representados pelos nomes pelos quais são chamados no software WEKA, de acordo com a tabela 4.1:

Tabela 4.1 – Nomes dos algoritmos para referência

Algoritmo	Nome WEKA
Regressão Logística	Logistic
Árvore de decisão J48	J48
Árvore de decisão RepTree	RepTree
Redes neurais	Multilayer Perceptron

4.1.1 Acurácia dos modelos

Frequentemente, adota-se o percentual de acertos, também chamado de acurácia do modelo, ao se aplicar um modelo de classificação (BARTH, 2004). As acurácias dos modelos são calculadas a partir de suas matrizes de confusão (FAWCETT, 2006). A matriz de confusão compara as classes reais com as previsões realizadas sobre as instâncias da base classificada. No caso deste estudo, como trata-se de um classificador binário, obtém-se uma matriz 2x2, como no exemplo da tabela 4.2, onde 0 representa as classificações de alunos adimplentes e 1 os alunos inadimplentes.

Tabela 4.2 - Exemplo de matriz de confusão

Real	Previsão			Total
	0	1	Total	
0	30	10	40	
1	5	55	60	
Total	35	65	100	

Neste exemplo, a acurácia do modelo = $(30+55) / 100 = 85\%$.

Além disso, temos que: (i) das 40 instâncias negativas, 30 foram corretamente classificadas (verdadeiro negativo), mas 10 foram incorretamente classificadas como positivo (falso positivo), erro conhecido como tipo II; (ii) e, das 60 instâncias positivas, tem-se 55 corretamente classificadas (verdadeiro positivo) e 5 erros (falso negativo), chamados de erro tipo I.

Além da taxa de acerto ou acurácia do modelo, outras estatísticas podem ser inferidas a partir da análise da matriz de confusão. Abaixo segue um breve descritivo das medidas que serão utilizadas neste estudo para fins comparativos.

Sensibilidade: também conhecido como recall ou taxa de verdadeiro positivo, mede o percentual de previsões de positivos, quando a classe real é positiva. No exemplo acima, $55/60 = 91.67\%$.

Especificidade: também conhecido como taxa de verdadeiro negativo, mede o percentual de previsões de negativo quando a classe real é negativa; $30/40 = 75\%$.

Precisão: é a taxa de acertos quando a previsão é positiva; $55/65 = 84.62\%$.

Na tabela 4.3 são dispostas as matrizes de confusão de cada modelo, obtidos pelo método de validação cruzada 10 vezes. Aqui, “inadimplente” representa a classe positiva.

Tabela 4.3 - Matrizes de confusão dos modelos calculados

Regressão Logística

Real	Previsão			Total
		Adimplente	Inadimplente	
Adimplente	1582	334	1916	
Inadimplente	600	2462	3062	
Total	2182	2796	4978	

Árvore de Decisão J48

Real	Previsão			Total
		Adimplente	Inadimplente	
Adimplente	1369	547	1916	
Inadimplente	587	2475	3062	
Total	1956	3022	4978	

RepTree

Real	Previsão			Total
		Adimplente	Inadimplente	
Adimplente	1508	408	1916	
Inadimplente	564	2498	3062	
Total	2072	2906	4978	

Redes Neurais

Real	Previsão			Total
		Adimplente	Inadimplente	
Adimplente	1365	551	1916	
Inadimplente	592	2470	3062	
Total	1957	3021	4978	

A tabela 4.4 resume a acurácia, sensibilidade, especificidade e previsão dos modelos calculados.

Tabela 4.4 - Resumo dos resultados de desempenho dos modelos

	Logit	J48	RepTree	Redes Neurais	Média
Acurácia	81,24%	77,21%	80,47%	77,04%	78,99%
Sensibilidade	80,40%	80,83%	81,58%	80,67%	80,87%
Especificidade	82,57%	71,45%	78,71%	71,24%	75,99%
Precisão	88,05%	81,90%	85,96%	81,76%	84,42%
Inadimplente/total	56,17%	60,71%	58,38%	60,69%	58,98%

Os resultados apresentados nas tabelas 4.3 e 4.4 são referentes às bases de teste obtidas via *10-fold cross-validation*, o que permite mais robustez e capacidade de generalização aos mesmos. Todos os quatro modelos utilizados neste trabalho, em média, apresentaram altos índices de acurácia, evidenciando a alta capacidade de prever o problema proposto. A regressão logística obteve o mais alto índice de acurácia dentre os algoritmos testados, e o método de redes neurais o mais baixo. Além disso, a regressão logística também demonstrou melhor desempenho na capacidade de prever corretamente a classe de inadimplência, com precisão de 88.05%, e o método de redes neurais, novamente, foi o que obteve mais baixa precisão. As taxas relativamente altas de sensibilidade e especificidade mostram, de modo geral, baixo risco de se cometer erros do tipo I e II. O risco de um falso negativo, ou seja, de classificar erradamente um aluno inadimplente como adimplente é de, em média, 19,13%, enquanto o risco de classificar erradamente como inadimplente um aluno adimplente (falso positivo) é cerca de 24.01%. Os altos índices de sensibilidade e especificidade demonstram que os atributos utilizados neste estudo são suficientes para identificar o aluno com risco de inadimplência.

É importante também comparar a acurácia média apresentada pelo uso dos modelos de *machine learning* com um modelo ingênuo (*naive*), que faz o uso da aleatoriedade nas previsões. Nestes modelos, a chance de um determinado aluno vir a se tornar inadimplente é de 62% (o número de alunos inadimplentes no total amostral), enquanto nos modelos que usam técnicas de previsão, esse percentual é de, em média, 79%. Isto significa que o uso de modelos de classificação permite um acerto em torno de 30% maior do que os modelos ingênuos, evidenciando a importância de uso de modelos estatísticos, como os utilizados neste estudo, para a análise de problemas gerenciais, como a inadimplência.

A prevalência da classe “inadimplente” na amostra – o número de alunos que estão inadimplentes sobre o total amostral – é de 62%. Todos os modelos exibiram um percentual menor da classe do que sua verdadeira prevalência. O algoritmo de J48 foi o que chegou mais perto em termos quantitativos, no entanto, considerando a sua baixa precisão se comparado aos outros métodos utilizados, pode-se afirmar que alcançou esse número mais alto classificando mais instâncias erradas. A alta prevalência da classe “inadimplente” é outro ponto interessante acerca desta análise, que demonstra a importância de se estudar a inadimplência como um

desafio gerencial, dado que, tanto nas previsões, como na realidade, a proporção de alunos inadimplentes é consideravelmente maior do que aqueles adimplentes.

4.1.2 Análise do modelo de regressão logística

O método da regressão logística, cuja origem data do início do século XIX, é hoje um dos mais populares para classificação binária (WILSON; LORENTZ, 2015). De acordo com Gonçalves, Gouvêa e Mantovani (2013), a regressão logística prediz as chances (*odds ratio*) de um evento binário - a classe - ocorrer. Como trata-se de uma razão entre probabilidades e estas oscilam entre 0 e 1, os possíveis valores de uma *odds ratio* variam de 0 a infinito positivo. Uma *odds ratio* entre 0 e 1 indica que aquele atributo está associado a uma menor chance de ocorrência da variável dependente. Uma razão igual 1 significa que o atributo não tem influência sobre a classe - o mesmo pode ser interpretado para números muito próximos de 1. Já valores maiores do que 1 indicam que o atributo aumenta as chances de ocorrência da classe e, quanto maior o valor, maior a chance (PEREIRA, 2017). A tabela 4.5 lista os atributos e suas respectivas *odds ratios*.

Tabela 4.5 - Atributos e *odds ratios* estimados pelo modelo de regressão logística

Atributos	Odds Ratios
Matrícula	1
Varição_média_valor_pago	1.4556
Média_valor_pago	1.0002
Média_qtde_disciplinas	4.1826
Varição_qtde_disciplinas	1.7391
Média_falta_por_disciplina	0.9139
Média_qtde_reprovações_por_falta	0.5628
Média_qtde_reprovações_por_nota	0.4727
Percentual_reprovações	0.0576
Média_reprovações	0.5537
Percentual_trancamentos	0.2586
Média_trancamentos	0.1474
Percentual_aprovações	0.2229
Média_aprovações	0.213
Qtde_débitos_18_1	0.4851
Qtde_débitos_17_2	0.7642
Qtde_débitos_17_1	0.8972

Sexo=M	0.9708
Idade	0.9983
Quantas_pessoas_moram_com_aluno	0.9818
Participação_vida_econômica_=_você_não_trabalha_e_seus_gastos_são_custeados=1	0.9861
Participação_vida_econômica_=_você_trabalha_e_é_responsável=1	0.8873
Participação_vida_econômica_=_você_trabalha_mas_não_é_independente=1	0.8635
Participação_vida_econômica_=_você_trabalha_e_é_independente_financeiramente=1	0.8029
Atividade_remunerada_=_emprego_autônomo=1	1.1397
Atividade_remunerada_=_sem_atividade=1	1.4919
Atividade_remunerada_=_nenhuma_das_opções_anteriores=1	0.9787
Atividade_remunerada_=_estágio=1	1.2089
Atividade_remunerada_=_emprego_fixo_particular=1	1.3064
Atividade_remunerada_=_emprego_fixo_federal_estadual_municipal=1	1.512
Contribuição_renda_familiar_=_pai=1	1.916
Contribuição_renda_familiar_=_mãe=1	1.7026
Contribuição_renda_familiar_=_cônjuge_ou_companheiro=1	1.5106
Contribuição_renda_familiar_=_você_mesmo=1	1.144
Contribuição_renda_familiar_=_outra_pessoa=1	1.721
Filhos_=_não=1	0.7874
Renda_familiar_=_acima_de_04_salários_mínimos=1	1.3001
Renda_familiar_=_até_02_salários_mínimos=1	1.5621
Renda_familiar_=_02_a_04_salários_mínimos=1	1.4758
Ensino_médio_=_escola_particular_com_bolsa_integral=1	0.5266
Ensino_médio_=_escola_particular_sem_bolsa=1	0.4379
Ensino_médio_=_escola_particular_com_bolsa_parcial=1	0.5675
Ensino_médio_=_escola_pública=1	0.4531

É importante ressaltar que a interpretação dos resultados obtidos por esse modelo de regressão logística é válida apenas para a amostra analisada, e não necessariamente possui significados reais sobre toda a população, no caso, todos os alunos de uma IES. Entretanto, a partir dos resultados dispostos, é possível inferir algumas informações, ainda que com algum nível de incerteza. A primeira delas diz respeito a relação entre os atributos financeiros com a inadimplência. O atributo financeiro que apresentou maior relação com a inadimplência foi a variação média do valor pago de semestralidade ao longo dos quatro semestres analisados (*odds ratio* de 1,45). Isto quer dizer que um aluno com maior variação média no valor pago ao longo dos semestres analisados tem 1,45 vezes mais chances de ser inadimplente do que um aluno com menor variação. Isso demonstra a importância de se realizar um estudo de sensibilidade de preços para entender a elasticidade da

demanda e saber o limite que os valores de mensalidade podem ser corrigidos e/ou o percentual de desconto adequado para cada aluno, a fim de evitar que ele venha a se tornar inadimplente por não conseguir arcar com as mensalidades.

Uma variação positiva de valor pago, isto é, o aumento do valor da semestralidade ao longo dos semestres analisados e o risco de inadimplência pode ser facilmente explicado pela forte relação entre a inadimplência e fatores financeiros, que afetam a capacidade pagadora dos alunos (quanto maior o valor de mensalidade pago por um aluno, maior o seu risco de inadimplência). Ademais, a IES analisada tem a política de conceder descontos na primeira semestralidade de seus cursos para atrair novos alunos, e ajustar esses valores ao longo do curso e, anualmente, há um reajuste nos valores de mensalidade. Estes aspectos podem influenciar na relação entre a inadimplência e a variação positiva de valor pago. Já a variação negativa do valor pago, isto é, a redução do valor pago ao longo do tempo, pode estar relacionado com a inadimplência à medida que pode ser um indicativo de problemas financeiros, onde o aluno procura reduzir a quantidade de disciplinas cursadas, para diminuir o valor da sua mensalidade e tentar se manter adimplente.

A média de disciplinas cursadas pelo aluno possui *odds ratio* igual a 4,18, o que indica uma forte relação com o risco de inadimplência (quanto mais disciplinas cursadas em média, maior a chance de inadimplência). A variação média de disciplinas cursadas ao longo dos semestres analisados também se apresentou relevante para o modelo, com *odds ratio* de 1,73. A análise dessas duas variáveis está relacionada com o atributo financeiro de variação do valor pago, avaliado anteriormente. Uma variação no número de disciplinas cursadas, afeta o valor pago pelo aluno, uma vez que o sistema de pagamento da IES é calculado de acordo com o número de disciplinas em que o aluno está matriculado. Tal fato deve ser motivo de atenção para a gestão das instituições, a fim de evitar que os alunos, almejando terminar logo os seus cursos, se matriculem em um número muito elevado de disciplinas a cada semestre e acabem se tornando inadimplentes por não conseguirem arcar com o valor das mensalidades ou por não conseguirem acompanhar academicamente as disciplinas, obtendo resultados insatisfatórios e se desmotivando com o curso.

As variáveis acadêmicas relacionadas às aprovações e reprovações médias por disciplina, também merecem ser analisadas. Com relação às variáveis de

aprovação, tanto a média de aprovações, como também o percentual, apresentaram *odds ratios* bem baixos, 0,21 e 0,22, respectivamente. Isto significa que quanto mais aprovações um aluno possui, menos chance de ele ser inadimplente, fato facilmente compreensível, uma vez que demonstra engajamento e interesse do aluno para com a instituição. Entretanto, para as variáveis de reprovações, o modelo também apresentou *odds ratios* baixas (média de reprovações 0,55 e percentual de reprovações 0,05), demonstrando que quanto mais reprovações um aluno possui, mais difícil de ele ser inadimplente. Uma possível explicação seria o fato que o aluno em uma situação de baixo rendimento acadêmico, acaba se desmotivando para com o curso e trancando ou cancelando sua matrícula antes de se tornar inadimplente. Tal explicação vai de encontro ao fato que as variáveis relacionadas à média e percentual de trancamentos apresentaram, respectivamente, *odds ratios* de 0,14 e 0,25, o que indica que quanto mais trancamentos o aluno realiza em um semestre, menor a sua chance de se tornar inadimplente, ou seja, para evitar uma situação de inadimplência, os alunos optam por trancar as disciplinas que estão matriculados.

Outro ponto interessante que vale ressaltar na análise dos resultados do modelo de regressão logística é a forte relação entre a inadimplência e a fonte de contribuição da renda familiar. Nos casos em que a renda da família provém do pai (1,91), mãe (1,7), cônjuge (1,51) ou outra pessoa (1,72), o risco de inadimplência é maior, se comparado aos casos em que o próprio aluno é o seu responsável financeiro (1,14). Isto pode ser explicado pelo fato de que, em um contexto onde o aluno não é independente financeiramente, as prioridades familiares, principalmente em momentos de crise econômica, acabam sendo divergentes e o comprometimento com a quitação das mensalidades não é o mesmo. Com relação a renda familiar, conforme já se esperava, o risco de inadimplência é maior para o grupo com a menor faixa de renda dentre as opções analisadas, de até 2 salários mínimos (*odds ratio* de 1,56).

No que diz respeito à atividade profissional exercida, os *odds ratios* foram maiores nos casos em que o aluno não exerce nenhuma atividade remunerada (1,49), ou seja, seus familiares ou outras pessoas são responsáveis pelo pagamento das mensalidades, o que vai de encontro com a análise anterior, onde o risco de inadimplência é maior quando o aluno não é o seu responsável financeiro. Alunos que trabalham em empresas públicas também possuem maiores chances de se

tornarem inadimplentes (1,51). Uma possível causa pode estar associada com a recente crise econômica enfrentada pelo país e principalmente o estado do Rio de Janeiro, o que leva a atrasos de pagamentos à funcionários públicos.

As variáveis relacionadas aos débitos em semestres anteriores apresentaram *odds ratios* abaixo de um (0,48, 0,76 e 0,89), o que indica que um aluno que teve débitos anteriores ao semestre analisado, possui menos chance de ser inadimplente. Uma possível explicação seria o fato de o aluno que já teve débitos com a IES e conseguiu quitar e acertar a sua situação, aprendeu com o ocorrido a ponto de evitar que aconteça novamente. Pode estar associado às políticas bem-sucedidas de cobrança da instituição, que, além de conseguirem fazer com que os alunos quitem seus débitos junto a IES, conseguem, também, evitar que se torne uma situação recorrente.

4.1.3 Análise do modelo de árvore de decisão

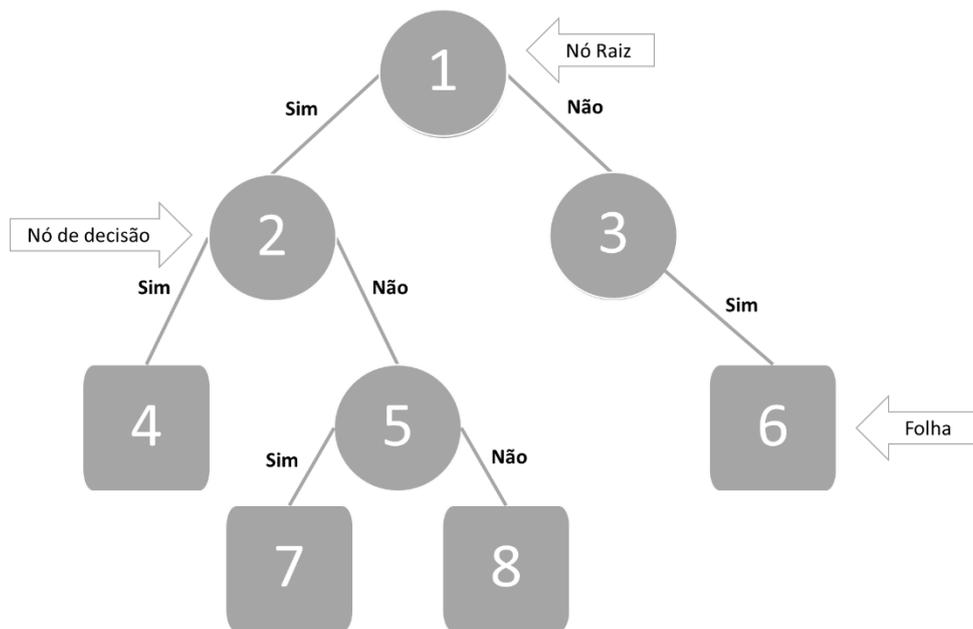
As árvores de decisão são funções que apresentam como dados de entrada um vetor de atributos e uma decisão como valor de saída: sim ou não (BREIMAN et al., 1984). Uma árvore de decisão utiliza a estratégia chamada dividir-para-conquistar, ou seja, um problema complexo é decomposto em subproblemas mais simples. Recursivamente, a mesma estratégia é aplicada a cada subproblema. A separação dos dados ocorre até que cada subconjunto esteja homogêneo, com casos de uma única classe (WITTEN et al., 2011).

O algoritmo de árvores de decisão é o único método de classificação de dados a apresentar os resultados hierarquicamente. Nelas, o atributo mais importante é apresentado na árvore como o primeiro nó, e os atributos menos relevantes são mostrados nos nós subsequentes (LEMOS; STEINER; NIEVOLA, 2005). Os nós representam regiões onde são realizados testes lógicos para a separação dos dados. O primeiro nó é chamado de nó raiz e é o nó principal da árvore de decisão. Os nós que estão localizados abaixo do nó raiz são os nós de decisão, que são divididos por ramos – essas divisões também são conhecidas como *splits*. As folhas são as regiões que estão associadas a um rótulo ou valor. Quando chega na folha, o dado é identificado como pertencente a uma das classes (BREIMAN et al., 1984).

De acordo com Lemos, Steiner e Nievola (2005), a principal vantagem das árvores de decisão é o elevado grau de compreensão de seus resultados, permitindo

uma tomada de decisões que leve em consideração os atributos mais relevantes. Ao escolher e apresentar os atributos em ordem de importância, as árvores de decisão permitem aos usuários conhecer quais fatores mais influenciam suas variáveis-alvo. A estrutura básica de uma árvore de decisão é ilustrada na figura 4.1.

Figura 4.1 - Estrutura geral de uma árvore de decisão



O tamanho total da árvore calculada pelo algoritmo J48 para este trabalho foi de 239 nós e folhas, e 120 níveis gerados. Dado o tamanho desta árvore, exibi-la junto de todos os seus nós pode ser complexo, mas examinar seus primeiros nós de decisão pode dar alguns *insights* para a gestão da instituição. Os primeiros *splits* indicam variáveis mais determinantes para a classificação da instância de acordo com o algoritmo, portanto, deve-se prestar mais atenção neles. No primeiro *split* aparece a quantidade de débitos em 2018.1, semestre anterior ao utilizado para a análise da classe, dividindo-se em dois ramos, débitos maior que 1 ou débitos menor ou igual a 1. Alguns resultados encontrados neste modelo corroboram aqueles encontrados no modelo de regressão logística, como a relevância da natureza do emprego do aluno (autônomo, sem emprego e empresa pública apresentaram maior relação com o risco de inadimplência) e o estado civil.

A figura 4.2 apresenta os primeiros *splits* da árvore de decisão gerada.

com que o pagamento das mensalidades nas datas aprazadas não se torne uma prioridade.

A primeira regra que podemos perceber na análise dos primeiros *splits* deste modelo de árvore de decisão é que um aluno terá maiores chances de se tornar inadimplente se a quantidade de débitos em 2018.1 (semestre anterior ao analisado) for menor ou igual do que 1 (ou seja, 0 ou 1 débitos), seguido pela quantidade de débitos em 2017.2 (dois semestres anteriores ao analisado) maior do que 1 e ser um autônomo. Caso o aluno não possua um emprego autônomo, suas chances de inadimplência serão maiores se ele for divorciado. Isto pode estar associado ao fato que indivíduos divorciados têm de arcar com despesas de pensões para filhos e/ou cônjuges, e o pagamento das mensalidades não se torna uma prioridade.

Por outro lado, se a quantidade de débitos no semestre anterior for maior do que 3, o aluno terá maiores chances de se tornar inadimplente. Esta folha foi a que apresentou maiores casos de inadimplência classificados corretamente, foram 1.155, representando uma probabilidade associada de 34.61% de chance de se tornar inadimplente. Caso a quantidade de débitos seja 2 ou 3, ele será inadimplente caso trabalhe em uma empresa pública. Se o aluno não trabalhar em uma empresa pública, mas for divorciado, também terá chances de se tornar inadimplente.

Ademais, outras variáveis apareceram no modelo (nos níveis mais avançados da árvore que não foram apresentados na figura 4.2) e merecem ser analisadas. A primeira delas é a variação média do valor pago. Se essa variação apresentar uma tendência positiva, isto é, um aumento da semestralidade ao longo do tempo, o aluno terá mais chances de se tornar inadimplente. Tal fato vai de encontro com a forte relação entre a inadimplência e questões financeiras, como aumentos de mensalidade, que afetam a capacidade pagadora dos alunos. Outro aspecto relevante apresentado pelo modelo é que se a quantidade média de disciplinas cursadas em um semestre for maior do que 6.25, o aluno terá mais chances de se tornar inadimplente, o que também pode ser explicado pelo fato que, quanto mais disciplinas um aluno cursa, maior o valor da sua mensalidade e, conseqüentemente, maior o risco de se tornar inadimplente.

Por outro lado, o modelo demonstrou que se a variação da quantidade de disciplinas cursadas for negativa, isto é, uma redução da quantidade de disciplinas ao longo do tempo, maior o risco de inadimplência. Uma possível explicação para isso seria o fato que o aluno já com dificuldades financeiras, começa a reduzir o

número de disciplinas cursadas para diminuir o valor da sua mensalidade e tentar não se tornar inadimplente.

Outro aspecto interessante exposto pelo modelo de árvore de decisão foi o fato de alunos que apresentaram percentual médio de trancamento de disciplinas menor do que 5%, têm mais chances de se tornarem inadimplentes. Além disso, o modelo evidenciou uma relação entre quantidades médias elevadas de reprovações tanto por nota, como por falta, e a inadimplência. Estes dois aspectos podem estar associados ao fato que o aluno inadimplente, comprovadamente desmotivado com o curso, não tem o hábito de trancar disciplinas para evitar reprovações.

4.1.4 Análise do modelo de redes neurais

As redes neurais artificiais (RNA) são modelos que buscam simular o processamento de informação do cérebro humano. São compostas por unidades de processamentos simples, os neurônios, que se unem por meio de conexões sinápticas (FERNEDA, 2006). Esses elementos são os responsáveis pela captação, computação e transmissão de informações e é através da atividade destes nodos que a rede tem a capacidade de assimilar dados fornecidos do ambiente externo e transformá-los em um vetor de resposta (HAYKIN, 2007).

A arquitetura de uma rede neural restringe o tipo de problema no qual a rede poderá ser utilizada, e é definida pelo número de camadas (camada única ou múltiplas camadas), pelo número de nós em cada camada, pelo tipo de conexão entre os nós (*feedforward* ou *feedback*) e por sua topologia (HAYKIN, 2007).

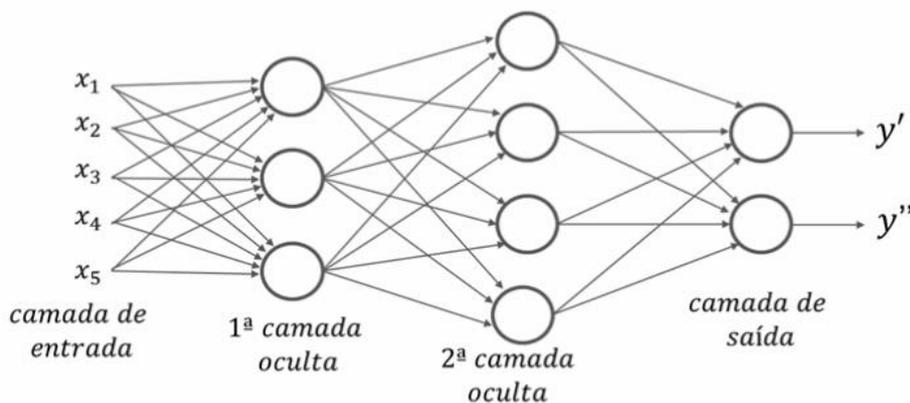
As RNAs são capazes de reconhecer padrões, extrair regularidades e detectar relações subjacentes em um conjunto de dados aparentemente desconexos. Neste sentido, sua capacidade de prever sistemas não lineares torna possível a sua aplicação em problemas de previsão de inadimplência (CAPANEMA E NASCIMENTO, 2016).

A aplicação de redes neurais constitui-se de uma abordagem relativamente nova, principalmente em estudos conduzidos no Brasil e, aos poucos, vem sendo introduzida no contexto de finanças, como nos estudos conduzidos por Mueller (1996), Arraes, Semolini e Picinini (1999), Neto, Nagano e Moraes (2006), Oliveira (2007), Lima et al. (2009) e Capanema e Nascimento (2016). Sobrinho (2017)

utilizou o algoritmo para avaliação da inadimplência em uma instituição de ensino básico.

De uma forma simplificada, uma rede neural artificial pode ser vista como um grafo onde os nós são os neurônios e as ligações fazem a função das sinapses, como exemplificado na figura 4.3.

Figura 4.3 - Estrutura geral de uma rede neural.
Fonte: Ferreira et al., 2016



O modelo de RNA gerado neste estudo, apesar de apresentar um índice de acurácia relativamente alto (77.04%), o que evidencia seu potencial preditivo, não apresenta explicações sobre as relações entre as variáveis deste problema. Não existe uma forma clara de entender a relação dos *outputs* do modelo com a variável-alvo estudada. Dessa forma, neste estudo, o modelo de RNA foi utilizado para uma análise comparativa entre os modelos propostos a fim de determinar aquele que possui a maior capacidade de prever o problema de inadimplência proposto.

4.2. Implicações do uso dos modelos

O objetivo do uso de modelos de *machine learning* neste estudo foi entender melhor os alunos de uma IES e os fatores que determinam a sua capacidade pagadora. Dentre os modelos propostos, a regressão logística foi a que apresentou melhor índice de acurácia e precisão, indicando a maior capacidade deste modelo em prever alunos com risco de inadimplência para os dados analisados. Sendo assim, o uso de tal modelo deveria ser incorporado pela IES nas ações futuras a serem adotadas contra a inadimplência.

Como passos futuros para a continuação dos estudos iniciados neste trabalho, propõe-se que o modelo de regressão logística seja aplicado em bases novas de alunos, para que estes possam ser classificados de acordo com o risco de inadimplência, além de, quando possível, estimar a probabilidade, ou *score*, de cada aluno vir a se tornar inadimplente. Estas probabilidades podem ser extraídas de alguns modelos de *machine learning*, utilizando o software de análise estatística SAS. Com isso, os modelos de previsão, ao analisarem um novo candidato a crédito, produzem um resultado numérico (*score*) que gerará uma classificação, como adimplente ou inadimplente, amparada pela definição de um ponto de corte, de acordo com as regras de cada negócio. Abaixo do ponto de corte, o candidato será considerado propenso à inadimplência (BARTH, 2004).

Além disso, sugere-se também, como resultado das análises dos modelos utilizados, a adoção pela IES de ações preventivas de natureza pedagógica, acadêmica e financeira contra a inadimplência. Na análise do modelo de regressão logística, por exemplo, os alunos mais propensos a se tornarem inadimplentes são aqueles que apresentam variações do valor de semestralidade, o que permite ações financeiras de descontos a fim de evitar que o aluno venha a se tornar inadimplente por não conseguir arcar com os valores de mensalidade. A quantidade de disciplinas também foi um fator preditivo da inadimplência neste modelo, o que permite uma interferência acadêmica na quantidade de disciplinas que um aluno pode vir a se matricular a cada semestre. O aluno não deve se afastar muito da quantidade média de disciplina que realiza ao longo dos períodos, e, para cada aluno, deve haver um limite de disciplinas a serem cursadas por semestre, de acordo com o seu histórico.

Considerando os resultados obtidos no modelo de árvore de decisão J48, as ações preventivas a serem tomadas dizem respeito, principalmente, a questão do devedor contumaz, uma vez que o principal resultado obtido pelo modelo relaciona fortemente a inadimplência com a existência de débitos nos semestres anteriores. Com isso, cabe a IES, por exemplo, o endurecimento de ofertas comerciais para aqueles devedores contumazes com a exigência de pagamentos maiores à vista, uma política de cobrança que entre em contato com esses alunos logo no primeiro débito com a IES, além de um reforço acadêmico para que esses alunos não se sintam desmotivados por estarem com muitos débitos com a IES e abandonem seus cursos.

A tabela 4.6 sintetiza as principais variáveis financeiras, acadêmicas e sociodemográficas relacionadas ao risco de inadimplência de um aluno, obtidas nos resultados dos modelos gerados.

Tabela 4.6 – Resumo dos principais resultados dos modelos

Variáveis	Financeiras	Acadêmicas	Sociodemográficas
Inadimplência	Variação valor pago de semestralidade	Média disciplinas cursadas	Pai, mãe, cônjuge ou outra pessoa responsável pela renda familiar
	Débitos em semestres anteriores (modelo J48)	Variação disciplinas cursadas	Renda familiar de até 2 salários mínimos
		Quantidade reprovações por falta e por nota (modelo J48)	Aluno com emprego autônomo, sem emprego ou empresa pública
		Média e percentual de trancamentos de disciplinas	Aluno com status "divorciado"

5 Considerações Finais

Este trabalho teve como objetivo propor e testar um modelo de *credit scoring* com o uso de técnicas de mineração de dados no setor de educação superior privada, utilizando dados financeiros, acadêmicos e sociais de alunos de uma IES privada do Rio de Janeiro. Para isso, realizou-se uma análise quantitativa, pautada em técnicas de *machine learning*, para verificar os fatores que mais influenciam este comportamento e estimar o risco de inadimplência de cada aluno.

Para se atingir o objetivo final do trabalho, foram percorridas as seguintes etapas: (i) coleta dos dados dos sistemas da IES, transformação das variáveis nos atributos relevantes para o problema e consolidação em um banco de dados próprio para leitura no software de *machine learning* WEKA; (ii) construção e teste de modelos preditivos de classificação no software WEKA a partir dos algoritmos de regressão logística, árvore de decisão (RepTree), J48 e redes neurais; (iii) análise comparativa dos resultados dos modelos calculados.

A relevância teórica deste estudo se encontra no fato de demonstrar, através da construção e aplicação de modelos e posterior avaliação dos resultados, a viabilidade do uso de técnicas de *machine learning* como uma ferramenta de previsão do risco de inadimplência de alunos, podendo ser adaptado para aplicação em outras instituições de ensino, evidenciando tratar-se de uma ferramenta de amplo espectro. Poucos estudos associaram o uso de técnicas de *machine learning* à previsão do risco de inadimplência de um aluno, entendendo os fatores mais relevantes que influenciam este comportamento e permitindo a realização de ações preventivas diferenciadas de acordo com o risco de cada aluno. Os estudos existentes não são direcionados ao setor de educação superior privada ou realizam uma análise qualitativa sobre a inadimplência, explorando rasamente os dados estatísticos.

Do ponto de vista gerencial, o estudo é de grande valia à medida que, diante da existência de uma legislação que protege os alunos que se tornam inadimplentes em relação a contraprestação dos serviços educacionais, e que limita os esforços

das IES em reaver esses pagamentos pendentes, os modelos de *credit scoring* propostos neste estudo permitem que as instituições possam amenizar os efeitos do risco de concessão de crédito educacional.

Ademais, os modelos propostos neste estudo permitem que as IES possam gerenciar seu relacionamento com os alunos, conhecendo os principais fatores que influenciam a sua capacidade pagadora. Um dos principais méritos deste estudo é a possibilidade de a IES identificar os alunos com potencial de risco de se tornarem inadimplentes e poder realizar ações preventivas direcionadas para estes alunos. Sem o uso de modelos que utilizam técnicas de *machine learning*, as ações são realizadas indiscriminadamente para todos os alunos. O uso dos modelos preditivos permite que a IES reduza e concentre os seus recursos no grupo de alunos com maior risco de inadimplência, obtendo resultados mais eficientes tanto no processo de cobrança, como no de retenção destes alunos. Outro aspecto interessante do uso dos modelos propostos neste estudo são a sua capacidade preditiva, que permite que a IES tenha conhecimento sobre o risco de um aluno se tornar inadimplente antecipadamente, possibilitando a realização de ações preventivas a fim de evitar que atinja um patamar irreversível.

O resultado da análise dos dados permite alguns *insights* para a gestão da IES. No modelo de regressão logística, o atributo de variação do valor pago apresentou forte relação com o risco de inadimplência, mostrando a importância de se realizar uma análise de sensibilidade antes de qualquer mudança de preços, sejam aumentos ou descontos. Como oportunidade de aprimoramento para novos estudos, levando em consideração os resultados obtidos neste, propõe-se o entendimento do nível ideal de mensalidade e possíveis descontos para cada aluno. Com isso, a IES se preserva do risco de conceder descontos além do necessário, ou de perder alunos que não conseguiram arcar com o valor das mensalidades e necessitavam de mais desconto.

A quantidade média de disciplinas realizadas também apareceu como um atributo bastante relevante para a inadimplência no modelo de regressão logística. Os alunos, com o objetivo de acabar seus cursos em um tempo mais rápido do que o previsto, se matriculam em um número maior de disciplinas do que conseguem acompanhar e acabam obtendo resultados acadêmicos insatisfatórios, levando a desmotivação com o curso e, conseqüentemente, com o pagamento das mensalidades. Este fato vai de encontro com a teoria apresentada por Lemos,

Ribeiro e Siqueira (2017), que apontam o rendimento acadêmico como uma das possíveis causas da inadimplência. Para estes autores, o fato do aluno apresentar baixo rendimento acadêmico poderá contribuir para o fortalecimento da falta de interesse em efetuar os pagamentos das mensalidades nas datas aprazadas, dando prioridade aos outros pagamentos que despertem maior interesse. Sendo assim, a IES pode prevenir tal comportamento através tanto do acompanhamento pedagógico do aluno para evitar que ele se desmotive por não conseguir acompanhar as disciplinas, como também através de uma ação acadêmica que limite a quantidade de disciplinas a serem cursadas a cada semestre.

Outro resultado obtido pelos modelos que merece atenção gerencial é na relação da inadimplência com a natureza da atividade profissional exercida pelo aluno e com a sua independência (ou não) financeira. No primeiro caso, os modelos mostraram uma maior relação entre alunos com empregos autônomos, públicos e sem emprego e a inadimplência. Para estes casos, a IES deve ter uma atenção especial e, se for necessário, já oferecer descontos para se antecipar e prevenir que o aluno venha a se tornar inadimplente. Com relação às questões de responsabilidade financeira, especificamente o modelo de regressão logística indicou que aqueles alunos em que os pais, cônjuges ou outras pessoas são responsáveis pelo pagamento das mensalidades estão mais propensos a se tornarem inadimplentes, se comparados com os casos em que o próprio aluno é seu responsável financeiro. Nestes casos, também cabe a IES o monitoramento e a adoção de ações financeiras preventivas que diminuam o possível risco de inadimplência para este grupo de alunos.

Um último *insight* obtido pela análise dos resultados que é válido de uma aplicação gerencial é a relação entre inadimplência e alunos com débitos em semestres anteriores, explícito pelo modelo de árvore de decisão J48. Estes alunos podem ser considerados como devedores contumazes, um comportamento motivado pela disparidade entre o que deseja e o que se pode pagar e/ou por uma desorganização financeira. Nestes casos, a IES deve endurecer as ofertas comerciais para estes alunos, através, por exemplo, da exigência de um valor maior à vista para que seja possível a renovação de matrícula para o semestre subsequente, ou a não oferta de boas condições de negociação para estes alunos, evitando, assim, que acumulem dívidas. Além disso, a IES deve estabelecer uma política de cobrança bem estruturada para esse perfil de alunos, de forma que recebam o contato da

instituição logo na primeira dívida, evitando também o acúmulo. Do ponto-de-vista acadêmico, cabe um acompanhamento tanto acadêmico, como também pedagógico para esses alunos, a fim de evitar que eles se sintam desmotivados por não conseguirem se manter adimplentes com a instituição e venham a evadir.

5.1 Oportunidades para futuras aplicações

O uso de técnicas de *machine learning* no setor de educação superior vem crescendo cada vez mais, e por tratar-se de uma ferramenta absolutamente escalável, permite aplicação em outras áreas do negócio, além do problema da inadimplência. Na área acadêmica, por exemplo, o *learning analytics* permite a individualização da aprendizagem de acordo com o perfil de cada aluno.

Como ampliação dos resultados obtidos por esse trabalho, propõe-se a criação de um algoritmo que, para qualquer base de alunos (desconhecida) inserida, consiga calcular o risco de cada aluno se tornar inadimplente. Seria uma ferramenta gerencial de apoio a decisão para qualquer IES, e não exclusiva para a instituição analisada neste estudo. Além disso, para o caso de novos alunos, esse modelo calcularia o percentual ideal de desconto, baseado em dados históricos de alunos com perfil semelhante.

A classificação de alunos como inadimplentes neste trabalho levou em consideração o aspecto de possuir um ou mais débitos de mensalidade. Como oportunidade para futuras aplicações, sugere-se o aprimoramento desta ferramenta para que, além da questão financeira, leve em consideração também outros aspectos na classificação de risco de inadimplência. Questões relacionadas ao desempenho acadêmico do aluno, e sociais, como a capacidade de influenciar pessoas e trazer novos alunos, em conjunto com os aspectos financeiros, podem ser relevantes e enriquecer o cálculo do risco de inadimplência de um aluno. Por exemplo, as ações preventivas acerca da inadimplência podem ser balizadas pelo fato de tratar-se, ou não, de um aluno com elevado desempenho acadêmico, ou pelo fato de ser um aluno com grandes chances de realizar indicações de novos alunos e boca-a-boca positivo da IES. Assim, a classificação de risco de inadimplência levaria em conta, também, aspectos de desempenho acadêmico e sociais dos alunos.

5.2 Limitações do Estudo

Este estudo leva em consideração aspectos financeiros, acadêmicos e sociais dos alunos de uma IES privada, desconsiderando a relação dos fatores externos, como mudanças econômicas e nível de desemprego no cálculo do risco de inadimplência. Além disso, como a base de dados utilizada é de apenas uma IES, não é possível avaliar as características da própria instituição que contribuíram para a inadimplência. Outra limitação da metodologia utilizada é a impossibilidade de se prever exatamente em que momento a inadimplência ocorrerá, uma vez que a análise não engloba um estudo temporal que forneça informações a respeito de em quanto tempo um aluno se tornará inadimplente. Por fim, a base de dados utilizada no estudo englobou somente alunos que estavam ativos nos quatro semestres dos anos de 2017 e 2018, podendo-se ampliar a série temporal para abranger períodos mais longos.

6 Referencial Bibliográfico

AHMAD, Rizal; BUTTLE, Francis. Customer retention: a potentially potent marketing management strategy. **Journal of strategic marketing**, v. 9, n. 1, p. 29-45, 2001.

AKKOÇ, Soner. An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for *credit scoring* analysis: The case of Turkish credit card data. **European Journal of Operational Research**, v. 222, n. 1, p. 168-178, 2012.

ALVES, César Moreira; CAMARGOS, Marcos Antonio. Fatores condicionantes da inadimplência em operações de microcrédito. **Revista Base (administração e Contabilidade) da UNISINOS**, v. 11, n. 1, p. 59-74, 2014.

ANDRADE, Simone Ferreira Capriccio et al. A inadimplência nas instituições particulares de ensino na cidade de Franca. **FACEF Pesquisa-Desenvolvimento e Gestão**, v. 11, n. 1, 2008.

ANNIBAL, C. A. Inadimplência do setor bancário brasileiro: uma avaliação de suas medidas. *Trabalhos para Discussão*, Brasília, Banco Central do Brasil, n. 192, p. 1-36, 2009.

ARAÚJO, Elaine Aparecida; CARMONA, Charles Ulises. Desenvolvimento de modelos *Credit scoring* com abordagem de regressão logística para a gestão da inadimplência de uma instituição de microcrédito. **Contabilidade Vista & Revista**, v. 18, n. 3, p. 107-131, 2007.

ARRAES, Daniel; SEMOLINI, Robinson; PICININI, Ronaldo. Arquiteturas de Redes Neurais Aplicadas a Data Mining no Mercado Financeiro. Uma Aplicação para a Geração de Credit Ratings. In: **IV Brazilian Congress of Neural Nets**, 1999.

- ASTIN, Alexander W. et al. **College student retention: Formula for student success**. 2 ed. Maryland: Rowman & Littlefield Publishers, 2012.
- BALLOU, Donald P.; TAYI, Giri Kumar. Enhancing data quality in data warehouse environments. **Communications of the ACM**, v. 42, n. 1, p. 73-78, 1999.
- BARNES, J. B. **Secrets of Customer Relationship Management**. New York: McGraw-Hill, 2001.
- BARTH, Nelson L. **Inadimplência: Construção de Modelos de Previsão**. São Paulo: Nobel, 2004.
- BEAVER, William H. Financial ratios as predictors of failure. **Journal of accounting research**, p. 71-111, 1966.
- BEM, Juliana Brugnola et al. **Um estudo sobre os impactos causados pela inadimplência e os reflexos na saúde financeira das empresas**. São Paulo, 2007. 78f. Monografia (Especialização em Administração) - Faculdade de Ciências Econômicas e Administrativas de Presidente Prudente.
- BERGAMO, Fábio Vinicius et al. A lealdade do estudante baseada na qualidade do relacionamento: uma análise em instituições de ensino superior. **BBR-Brazilian Business Review**, v. 9, n. 2, p. 26-47, 2012.
- BERGAMO, Fábio Vinicius et al. Modelo de lealdade e retenção de alunos para instituições do ensino superior: um estudo teórico com base no marketing de relacionamento. **BBR-Brazilian Business Review**, v. 8, n. 2, p. 43-66, 2011.
- BERGAMO, Fábio Vinicius; FARAH, Osvaldo Elias; GIULIANI, Antônio Carlos. A lealdade e a educação superior: ferramenta estratégica para a retenção de clientes. **Revista Ibero Americana de Estratégia**, v. 6, n. 1, p. 55-62, 2007.
- BERRY, M.; LINOFF, G. Mastering Data Mining: The Art and Science of Customer Relationship Management. **Industrial Management & Data Systems**, v. 100 n. 5, p. 245-246, 2000.
- BLATT, Adriano. **Avaliação de risco e decisão de crédito: um enfoque prático**. São Paulo: Nobel, 1999.

BRASIL. **Lei nº. 9.870, de 23 de novembro de 1999**. Dispõe sobre o valor total das anuidades escolares e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/19870.htm.

BREIMAN, L. et al. **Classification and Regression Trees**. Belmont: Wadsworth, 1984.

BRIGHAM, Eugene F.; GAPENSKI, Louis C.; EHRHARDT, Michael C. **Administração financeira – teoria e prática**. São Paulo: Atlas, 2001.

CAMARGOS, Marcos Antônio et al. Fatores condicionantes de inadimplência em processos de concessão de crédito a micro e pequenas empresas do estado de Minas Gerais. **Revista de Administração Contemporânea**, v. 14, n. 2, p. 333-352, 2010.

CAPANEMA, Daniel; NASCIMENTO, João Paulo Barbosa. Previsão de Inadimplência em uma Instituição de Ensino Superior Baseada em Redes Neurais Artificiais. **SYNTHESIS| Revistal Digital FAPAM**, v. 7, n. 7, p. 287-309, 2016.

CASA NOVA, Silvia Pereira de Castro. Quanto pior, melhor: Estudo da utilização da análise por envoltória de dados em modelos de análise de inadimplência/insolvência de empresas. **Revista Contemporânea de Contabilidade**, v. 10, n. 19, p. 71-96, 2013.

CECHINEL, Cristian; ARAUJO, Ricardo Matsumura; DETONI, Douglas. Modelagem e Predição de Reprovação de Acadêmicos de Cursos de Educação a Distância a partir da Contagem de Interações. **Revista Brasileira de Informática na Educação**, v. 23, n. 3, p. 1-11, 2015.

COLOMBO, Sonia Simões. **Marketing educacional em ação: estratégias e ferramentas**. Porto Alegre: Bookman, 2005.

CÔRTEZ, S. C.; PORCARO, R. M.; LIFSCHITZ, S. **Mineração de dados: funcionalidades, técnicas e abordagens**. 2002. 35 f. Monografia (Especialização em Informática) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.

COSTA, Oberdan Santos; GOUVEIA, Luis Borges. Modelos de retenção de estudantes: abordagens e perspectivas. **Revista Eletrônica de Administração**, v. 24, n. 3, p. 155-182, 2018.

CROSBY, Lawrence A.; EVANS, Kenneth R.; COWLES, Deborah. Relationship quality in services selling: an interpersonal influence perspective. **Journal of marketing**, v. 54, n. 3, p. 68-81, 1990.

FAWCETT, Tom. An introduction to ROC analysis. **Pattern recognition letters**, v. 27, n. 8, p. 861-874, 2006.

FENSTERSTOCK, Albert. Credit scoring and the next step. **Business credit**, v. 107, n. 3, p. 46-49, 2005.

FERNEDA, Edberto. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, v. 35, n. 1, p. 25-30, 2006.

FERREIRA, Arthur et al. Um estudo sobre previsão da demanda de encomendas utilizando uma rede neural artificial. **Blucher Marine Engineering Proceedings**, v. 2, n. 1, p. 353-364, 2016.

FREIRE, Adriana Cardoso; FREIRE, Flávia Magalhães. Causas da inadimplência nas instituições de ensino superior e meios de reduzi-la. **Renefara**, v. 5, n. 5, p. 247-259, 2014.

GARCIA, S.C. **O uso de árvores de decisão na descoberta de conhecimento na área da saúde**. 2000. 88f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal do Rio Grande do Sul, Porto Alegre.

GITMAN, Lawrence J. **Princípios de administração financeira - Essencial**. Porto Alegre: Bookman, 2001.

GONÇALVES, Eric Bacconi; GOUVÊA, Maria Aparecida; MANTOVANI, Daielly Melina Nassif. Análise de risco de crédito com o uso de regressão logística. **Revista Contemporânea de Contabilidade**, v. 10, n. 20, p. 139-160, 2013.

GOTTARDO, Ernani; KAESTNER, Celso; NORONHA, Robinson Vida. Avaliação de desempenho de estudantes em cursos de educação a distância utilizando mineração de dados. In: **Anais do Workshop de Desafios da Computação Aplicada à Educação**, v. 23, Rio de Janeiro, 2012.

GUJARATI, Damodar N.; PORTER, Dawn C. **Econometria Básica-5**. Porto Alegre: Amgh, 2011.

GUJARATI, Damodar. Use of dummy variables in testing for equality between sets of coefficients in two linear regressions: a note. **The American Statistician**, v. 24, n. 1, p. 50-52, 1970.

HALL, M., et al. The WEKA data mining software: an update. **ACM SIGKDD explorations newsletter**, 11.1, p. 10-18, 2009.

HAN, J.; KAMBER, M. **Data mining: Concepts and Techniques**. San Francisco: Morgan Kaufmann, 2006.

HAYKIN, Simon. **Redes neurais: princípios e prática**. Porto Alegre: Bookman, 2007.

HENNIG-THURAU, Thorsten; LANGER, Markus F.; HANSEN, Ursula. Modeling and managing student loyalty: An approach based on the concept of relationship quality. **Journal of service research**, v. 3, n. 4, p. 331-344, 2001.

HERLING, Luiz Henrique et al. A inadimplência nas instituições de ensino superior: um estudo de caso na instituição XZX. **Revista Gestão Universitária na América Latina-GUAL**, v. 6, n. 2, p. 126-142, 2013.

HOSMER, David W.; LEMESHOW, Stanley; STURDIVANT, Rodney X. Introduction to the logistic regression model. **Applied logistic regression**, v. 2, p. 1-30, 2000.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Censo da Educação Superior**. Brasília: INEP, 2018. Disponível em: <http://portal.inep.gov.br/web/centso-da-educacao-superior>. Acesso em: 19 set. 2019.

JANG, Jyh-Shing Roger; SUN, Chuen-Tsai; MIZUTANI, Eiji. Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]. **IEEE Transactions on automatic control**, v. 42, n. 10, p. 1482-1484, 1997.

KOVÁCS, Zsolt László. **Redes neurais artificiais**. São Paulo: Livraria da Física, 2002.

KUMAR, V. A Theory of Customer Valuation: Concepts, Metrics, Strategy, and Implementation. **Journal of Marketing**, v. 82, n.1, p. 1-19, 2018.

KUMAR, V. et al. Undervalued or Overvalued Customers: Capturing Total Customer Engagement Value. **Journal of Service Research**, v. 13, n.3, p. 297-310, 2010.

LEMOS, Anderson Queiroz; RIBEIRO, Fernando; SIQUEIRA, Elismar Oliveira. O Acesso Ao Ensino Superior E O Problema Da Inadimplência: Um Estudo Sobre Os Fatores Determinantes Da Capacidade Pagadora De Alunos Numa Instituição Particular. **Intern. Journal of Profess. Bus. Review**, v. 2, n. 2, p. 23-35, 2017.

LEMOS, Eliane Prezepiorski; STEINER, Maria Teresinha Arns; NIEVOLA, Julio César. Análise de crédito bancário por meio de redes neurais e árvores de decisão: uma aplicação simples de data mining. **Revista de Administração-RAUSP**, v. 40, n. 3, p. 225-234, 2005.

LEWIS, Edward M. **An introduction to credit scoring**. California: Fair, Isaac and Company, 1992.

LIMA, Fabiano Guasti et al. Aplicação de redes neurais na análise e na concessão de crédito ao consumidor. **Revista de Administração-RAUSP**, v. 44, n. 1, p. 34-45, 2009.

MACHADO, Dorival dos S. **Redução da Inadimplência no Setor da Educação**. São Paulo: SRS Editora, 2009.

MAIA, A. S. R. S. **Inadimplência e Recuperação de Créditos**. 2007. 65p. Trabalho de Conclusão de Curso (Especialização em Administração) – Universidade Federal do Rio Grande do Sul, Londrina.

MATIAS, A. B. (org.). **Finanças corporativas de curto prazo**. São Paulo: Atlas, 2007.

MCCLAVE, James T. et al. **Statistics for business and economics**. Boston: Pearson, 2014.

NETO, Sigismundo Bialoskorski; NAGANO, Marcelo Seido; MORAES, Marcelo Botelho. Utilização de redes neurais artificiais para avaliação socioeconômica: uma aplicação em cooperativas. **Revista de Administração-RAUSP**, v. 41, n. 1, p. 59-68, 2006.

OLIVEIRA, Mauri Aparecido de. **Aplicação de redes neurais artificiais na análise de séries temporais econômico-financeiras**. 2007. 316 f. Tese (Doutorado em Administração) - Universidade de São Paulo, São Paulo, 2007.

OLIVEIRA, Rodrigo Teles Dantas; BARBOSA, Enny Dantas. Retenção dos discentes de administração da UFS: fatores condicionantes e ações da

gestão acadêmica. **Administração: Ensino e Pesquisa**, v. 17, n. 2, p. 355, 2016.

PALMUTI, Claudio Silva. **Mensuração do risco de crédito através de análise estatística multivariada**. 2012. 89 f. Dissertação (Mestrado em Administração de Empresas) – Escola de Administração, FACCAMP, São Paulo.

PAULA, G. A. Modelos de Regressão com Apoio Computacional. Instituto de Matemática e Estatística, 2002. Disponível em: http://www.pucrs.br/ciencias/viali/tic_literatura/livros/texto_2013.pdf. Acesso em: 03 dez. 2019.

PEPPERS, D.; ROGERS, M. **Managing Customer Relationships: A Strategic Framework**. 2 ed. Nova Jersey: John Wiley & Sons, Inc, 2011.

PEREIRA, Francisco Coimbra Carneiro. **Modelos Preditivos para Evasão de Alunos no Ensino Superior Privado – Uma aplicação de Machine Learning para Gestão de Marketing de Relacionamento**. 2017. 69 f. Dissertação (Mestrado em Administração de Empresas) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.

PINHEIRO, Joana Caroline de Araújo. **Dano moral e a inclusão indevida nos bancos de dados e cadastros de inadimplentes: o cabimento de indenização ao devedor contumaz**. 2015. 61f. Monografia (Especialização em direito) - Universidade Federal do Maranhão, São Luís.

PYLE, D. **Data Preparation for Data Mining**. San Francisco: Morgan Kaufmann, 1999.

ROSENBERG, Eric; GLEIT, Alan. Quantitative methods in credit management: a survey. **Operations research**, v. 42, n. 4, p. 589-613, 1994.

ROSS, Stephen A. et al. **Administração financeira**. New York: AMGH Editora, 2015.

SANTOS, José Odálio. **Análise de crédito – empresas e pessoas físicas**. São Paulo: Atlas, 2003.

SANTOS, José Odálio; FAMÁ, Rubens. Avaliação da aplicabilidade de um modelo de *credit scoring* com variáveis sistêmicas e não-sistêmicas em carteiras de crédito bancário rotativo de pessoas físicas. **Revista Contabilidade & Finanças**, v. 18, n. 44, p. 105-117, 2007.

SAUNDERS, Anthony. **Medindo o risco de crédito: novas abordagens para value at risk e outros paradigmas**. Rio de Janeiro: Qualitymark, 2000.

SCHIMITT, P.; SKIERA, B.; VAN DEN BULTE, C. Referral Programs and Customer Value. **Journal of Marketing**, v. 75, p. 46-59, 2011.

SCHRICKEL, Wolfgang Kurt. **Análise de Crédito: Concessão e Gerência de Empréstimos**. São Paulo: Atlas S.A, 1998.

SHAJAHAN, S. **Relationship marketing: text and cases**. New Delhi: Tata McGraw-Hill Education, 2006.

SILVA FILHO, R. L. L. et al. **A evasão do ensino superior brasileiro**. Cadernos de Pesquisa, São Paulo, v. 37, n. 132, p. 641-659, 2007.

SILVA, José Pereira da. **Gestão e análise de risco de crédito**. São Paulo: Atlas, 1997.

SILVA, Pablo Rogers. **Psicologia do risco de crédito: análise da contribuição de variáveis psicológicas em modelos de credit scoring**. 2011. 244 f. Tese (Doutorado em Administração) - Universidade de São Paulo, São Paulo.

Sindicato das Entidades Mantenedoras dos Estabelecimentos de Ensino Superior no Estado de São Paulo – SEMESP. **Pesquisa de Inadimplência 2019**. São Paulo: SEMESP, 2019. Disponível em: <https://www.semesp.org.br/pesquisas/pesquisa-de-inadimplencia-2019/>. Acesso em: 04. set. 2019.

Sindicato das Entidades Mantenedoras dos Estabelecimentos de Ensino Superior no Estado do Rio de Janeiro – SEMERJ. **Panorama do Ensino Superior no Estado do Rio de Janeiro**. Rio de Janeiro: SEMERJ, 2019.

SOBRINHO, Marcelo José Vieira. **Um estudo da inadimplência aplicada ao segmento educacional de ensino médio e fundamental, utilizando modelos *credit scoring* com análise discriminante, regressão logística e redes neurais**. 2007. 138 f. Dissertação (Mestrado em Administração de Empresas) – Universidade Federal de Pernambuco, Recife.

TEIXEIRA, Adriano; SILVA, Alexandre Ricardo. **Créditos de difícil recebimento: crédito, cobrança, inadimplência e os seus tratamentos contábeis**. 1 ed. Franca: Facef, 2001.

THOMAS, Lyn C. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. **International journal of forecasting**, v. 16, n. 2, p. 149-172, 2000.

TINTO, Vincent. Dropout from higher education: A theoretical synthesis of recent research. **Review of educational research**, v. 45, n. 1, p. 89-125, 1975.

TINTO, Vincent. Taking retention seriously: Rethinking the first year of college. **NACADA journal**, v. 19, n. 2, p. 5-9, 1999.

VIANNA, Cleverson Tabajara et al. Ranking de universidades no Brasil: um estudo de predição com machine learning. **SUCEG-Seminário de Universidade Corporativa e Escolas de Governo**, v. 1, n. 1, p. 363-379, 2017.

VIEIRA, Marcos Villela. **Administração estratégica do capital de giro**. 2 ed. São Paulo: Atlas, 2005.

WILSON, J. R.; LORENZ, K. A. **Short History of the Logistic Regression Model**. In: Modeling Binary Correlated Responses using SAS, SPSS and R, p. 1723, Springer International Publishing, 2015.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data Mining: Practical Machine Learning Tools and Techniques**. 3 ed. Burlington: Morgan Kaufmann, 2011.