

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ICEX – Instituto de Ciências Exatas
Programa de Pós-Graduação em Estatística

Juliana Simões da Silva Dutra

**CLASSIFICAÇÃO DE CLIENTES INADIMPLENTES VIA REGRESSÃO
LOGÍSTICA**

Belo Horizonte
2021

Juliana Simões da Silva Dutra

**CLASSIFICAÇÃO DE CLIENTES INADIMPLENTES VIA REGRESSÃO
LOGÍSTICA**

Monografia de Especialização apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Especialista em Estatística.

Orientador: Marcelo Azevedo Costa

Coordenador: Roberto da Costa Quinino

Belo Horizonte
2021

2021, Juliana Simões da Silva Dutra.
Todos os direitos reservados

Dutra, Juliana Simões da Silva.

D978c Classificação de clientes inadimplentes via regressão
logística [manuscrito] / Juliana Simões da Silva Dutra.—
2021.
26.f. il.

Orientador: Marcelo Azevedo Costa.
Coorientador: Roberto da Costa Quinino.
Monografia (especialização) - Universidade Federal
de Minas Gerais, Instituto de Ciências Exatas,
Departamento de Estatística.
Referências: f. 26.

1. Estatística. 2. Cobrança de contas. 3. Devedores.
4. Probabilidades. I. Costa, Marcelo Azevedo. II.
Quinino, Roberto da Costa . III. Universidade Federal de
Minas Gerais, Instituto de Ciências Exatas,
Departamento de Estatística .IV. Título.

CDU 519.2 (043)

Ficha catalográfica elaborada pela bibliotecária Belkiz Inez Rezende Costa CRB 6ª
Região nº 1510




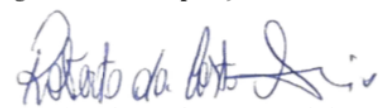
Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

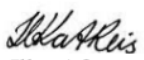
E-mail: pgest@ufmg.br
Tel: 3409-5923 – FAX: 3409-5924

ATA DO 233ª. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE JULIANA SIMÕES DA SILVA DUTRA.

Aos trinta dias do mês de agosto de 2021, às 16:30 horas, com utilização de recursos de videoconferência a distância, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso da aluna **Juliana Simões da Silva Dutra**, intitulado: “*Classificação de clientes inadimplentes via regressão logística*”, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, o Presidente da Comissão, Professor Marcelo Azevedo Costa – Orientador, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra à candidata para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa da candidata. Após a defesa, os membros da banca examinadora reuniram-se sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: a candidata foi considerada Aprovada condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 30 de agosto de 2021.


Prof. Marcelo Azevedo Costa (Orientador)
Departamento de Engenharia da Computação / Escola de Engenharia / UFMG


Prof. Roberto da Costa Quinino
Departamento de Estatística / ICEX / UFMG


Prof.ª Ilka Afonso Reis
Departamento de Estatística / ICEX / UFMG

AGRADECIMENTOS

Primeiramente agradeço a Deus pelo dom da vida e pela saúde para conseguir lutar por meus ideais.

Aos meus pais, Mariêta e Adão, pelo amor e dedicação.

Ao meu noivo, Felipe, que entendeu as ausências durante as horas de estudo e me apoiou nessa jornada.

Ao meu tio e padrinho Eli Lopes, agradeço o exemplo de pesquisador, de persistência e pelos ensinamentos durante a redação deste trabalho.

Ao escritório de advocacia onde trabalho, que cedeu os dados para análise e serviu de ambiente de pesquisa, em especial ao Diretor Thiago Peixoto, por ser um incentivador da cultura Data-Driven e ao Analista de Business Intelligence Douglas Henrique Perdigão, por me auxiliar com a geração da base de dados.

Agradeço aos professores, em especial ao meu orientador, Prof. Marcelo Azevedo Costa, por me acompanhar nesse momento tão importante que foi a construção do trabalho de conclusão de curso.

Enfim, sempre serei grata a todos que me incentivaram, que colaboraram com a minha formação e que hoje celebram comigo a concretização deste trabalho.

RESUMO

Diante da necessidade de entender o perfil do cliente inadimplente e criar estratégias de cobrança mais eficientes, o objetivo desse trabalho de conclusão de curso é classificar os clientes inadimplentes de uma base de cartões de crédito Private Label de acordo com a propensão de negociação de sua dívida. Os dados foram obtidos por meio do banco de dados de uma assessoria de cobrança de Belo Horizonte/MG, que presta serviço para uma grande instituição bancária de nível nacional. Foi ajustado um modelo de regressão logístico e criado um ranking de clientes, de acordo com sua propensão de negociação da dívida.

Palavras-Chave: atraso, cobrança, régua de cobrança, inadimplente, probabilidade.

ABSTRACT

Due to the necessity to understand the defaulting customers profile and create more efficient collection strategies, the objective of this study conclusion is to classify default by a private card company's customers according to the customer's propensity to negotiate debt. The data were extracted from a collection center's database located in Belo Horizonte, which provides services to a national large bank institution. A logistic regression model was adjusted and a ranking of clients was created according to their propensity to negotiate debt.

Keywords: delay, collection, collection ruler, defaulter, probability.

SUMÁRIO

1 INTRODUÇÃO	10
1.1 Problema.....	10
1.2 Objetivo	11
1.3 Justificativa.....	11
2.1 Delineamento	13
2.2 Aspectos éticos	13
2.3 Base da dados	14
2.4 Método	17
3 RESULTADOS E DISCUSSÕES.....	20
3.1 Caracterização da amostra	20
3.2 Ajuste do modelo de regressão logístico	23
4 CONCLUSÃO	27

LISTA DE GRÁFICOS

Gráfico 1: Valor das dívidas por fase (faixa de atraso)	20
Gráfico 2: Valor dos acordos por fase (faixa de atraso).....	21
Gráfico 3: Valor da parcela por fase (faixa de atraso).....	21
Gráfico 4: Desconto aplicado por fase (faixa de atraso)	22
Gráfico 5: Análise de correlação	22

LISTA DE FIGURAS

Figura 1: Curvas de sensibilidade e especificidade	25
Figura 2: Curva ROC.....	26

1 INTRODUÇÃO

Diante do atual cenário de pandemia e incertezas econômicas, os dados da inadimplência no Brasil mostram que o volume de consumidores endividados cresceu 2,7% comparando o primeiro semestre de 2020 ao mesmo período de 2019, enquanto em empresas o aumento foi de 7,3% (Serasa Experian, 2020).

Algumas medidas governamentais adotadas nesse período, tais como flexibilização em prazos de negativação e protestos de dívidas, condições especiais de negociação, entre outros, podem auxiliar no processo de renegociação e recuperação de crédito, impactando positivamente na retomada da economia.

É uma equação simples: investimento em cobrança + negociações adequadas = aumento da recuperação. A roda não pode parar de girar; com respeito, inteligência e ferramentas adequadas, sua empresa recupera dívidas e se prepara para os próximos ciclos econômicos, que sugerem ser mais promissores. (SERASA EXPERIAN, 2020).

Atuar de forma assertiva e ágil com a recuperação de crédito será um diferencial competitivo em empresas para melhorar sua fonte de receita, pois devido a pandemia, muitos setores tiveram suas atividades reduzidas, comprometendo faturamento e fluxo de caixa. Conhecer o perfil dos clientes e entender qual a melhor forma de abordá-lo, pode fazer diferença no momento da efetivação da negociação, garantindo sucesso na recuperação da dívida.

O cenário de pandemia fortaleceu ainda mais a importância dos canais digitais no processo de recuperação de crédito e a utilização de ferramentas que auxiliam no entendimento do perfil do devedor, possibilitando melhores estratégias de cobrança e criação de réguas específicas para cada caso, reduzindo atrito, prezando pela experiência do cliente e consequente retomada de um relacionamento com a instituição credora.

1.1 Problema

Como classificar os clientes inadimplentes de uma base de cartões de crédito Private Label de acordo com a propensão de negociação de sua dívida?

1.2 Objetivo

Classificar os clientes inadimplentes de uma base de cartões de crédito Private Label de acordo com a propensão de negociação de sua dívida, através de um modelo de regressão logístico, possibilitando melhorar as estratégias de cobrança, rentabilidade por canal de atendimento e ou segmentos.

1.3 Justificativa

O mercado de crédito é fundamental no cenário econômico do país, por estimular atividades financeiras, investimentos e aquisição de bens de consumo. Em contrapartida, a expansão do crédito pode contribuir para aumento da inadimplência e exposição das instituições credoras. Dessa forma, se faz necessário estudos e análises preditoras que sejam capazes de avaliar os riscos na concessão de empréstimos e na recuperação de dívidas.

Com o aumento da inadimplência, a área de cobrança das empresas ou a terceirização do processo, passam a ser pontos estratégicos na gestão dos negócios. Dentre os desafios a serem enfrentados, pode-se citar:

- Redução da PDD (Provisão Devedores Duvidosos);
- Criação de regras eficientes de priorização da cobrança (quem deve ser cobrado primeiro e quando);
- Mitigação de custos com o valor ao longo do tempo dos clientes;
- Implantação de ferramentas de negociação que permitam registros históricos para retroalimentação dos fluxos de cobrança.

Frente aos desafios estabelecidos, as empresas precisam evoluir seus modelos de maturidade de capacidade de cobrança (Deloitte, 2011). Quanto maior a capacidade de previsibilidade e impacto na receita, maior o diferencial competitivo das instituições.

- **Empresas emergentes:** possuem apenas uma célula de cobrança, processos operacionais manuais, ausência de indicadores, metas ou orçamentos operacionais.
- **Empresas avançadas:** automação da lista de cobrança, remuneração variável por resultados, equipe especializada em cobrança, relatórios baseados em estratégias de cobrança e células de cobrança segmentada.
- **Empresas líderes:** modelos de Collection Scoring, equipes especializadas, ferramentas de otimização e monitoramento avançados, workflow parametrizado e integrado para cobrança, scorecards de eficiência dos processos.

Dentre os principais modelos preditivos relacionados ao processo de cobrança pode-se citar o *Collection Scoring*¹, utilizado para avaliar a probabilidade de um cliente inadimplente se tornar adimplente. Esse modelo pode ser utilizado em diversas etapas do processo, de forma preventiva ou reativa.

- **Modelos preventivos:** criação de réguas para clientes adimplentes com grandes chances de se tornarem inadimplentes;
- **Modelos reativos:** criação de réguas durante o ciclo de cobrança, otimizando a alocação de recursos e priorização de canais, além de reduzir acordos desnecessários com grande chance de quebra.

Nesse contexto, a utilização de modelos preditivos pode melhorar a gestão da cobrança e aumentar efetivamente o retorno financeiro das instituições. As empresas precisam de métodos que possibilitem a gestão estratégica sobre os riscos que envolvem a recuperação de crédito e sua rentabilidade.

¹ Descrição mais detalhada do modelo de Collection Scoring pode ser obtida em Machado (2015).

2 MATERIAIS E MÉTODOS

2.1 Delineamento

Os inadimplentes que compõe a base de dados foram separados em dois grupos, de acordo com o status do seu pagamento. Dessa forma, fica definido como variável resposta o status de pagamento, sendo: 1- Bom pagador e 0- Mau pagador.

1- Bom pagador: negociou sua dívida e realizou o pagamento da primeira parcela;

0- Mau pagador: não negociou sua dívida ou negociou, mas não efetivou o pagamento da primeira parcela (quebra do acordo).

2.2 Aspectos éticos

A utilização de modelos preditivos para análise de riscos de crédito é resguardada pelo Acordo de Basiléia II, que permite a utilização de técnicas para avaliação de riscos aos quais as instituições estão sujeitas.

A Resolução nº 2682, de 21 de dezembro de 1999 do Banco Central do Brasil, que “dispõe sobre critérios de classificação das operações de crédito e regras para constituição de provisão para créditos de liquidação duvidosa” determina critérios que devem ser utilizados nesse processo.

A classificação da operação no nível de risco correspondente é de responsabilidade da instituição detentora do crédito e deve ser efetuada com base em critérios consistentes e verificáveis, amparada por informações internas e externas [...]. (BANCO CENTRAL DO BRASIL, 1999).

Além disso, uma preocupação atual das empresas é estar alinhada com as conformidades da Lei Geral de Proteção de Dados que inicia a cultura de proteção de dados no Brasil. As instituições financeiras e empresas de cobranças em geral, atuam com coleta, armazenamento e compartilhamento de informações pessoais de seus clientes.

A base de dados das instituições financeiras e as carteiras de clientes cedidas às empresas terceirizadas de cobrança possuem dados pessoais, sensíveis e

anonimizados, que devem ser tratados de acordo com as determinações da LGPD para segurança dos clientes.

- **Dados pessoais:** informações necessárias para identificar ou localizar uma pessoa;
- **Dados sensíveis:** características de cunho mais íntimo, relacionadas a questões biológicas, psicológicas, raciais, religiosas e sociais;
- **Dados anonimizados:** aspectos que não identificam direta ou indiretamente o indivíduo.

As empresas e/ou instituições financeiras podem ser controladoras ou operadoras dos dados e realizar tratamento deles, ou seja, armazenamento, acesso, utilização, reprodução, eliminação ou modificação. Algumas exigências da LGPD são fundamentais para que o tratamento dos dados nas operações de cobrança ocorra de forma adequada:

- Permissão para manipulação de dados objetivando a proteção do crédito;
- Os dados manipulados devem ter relação direta com a operação e modelo de negócio controladora dos dados;
- Utilização onde os dados sejam essenciais para a execução de contratos ou procedimentos relacionados;
- Armazenamento de dados para cumprir determinada exigência judicial ou normalização.

Além disso, cuidados com segurança da informação e governança de dados são fundamentais para cumprimento das diretrizes da LGPD e proteção dos dados dos titulares.

2.3 Base da dados

A base de dados utilizada neste trabalho foi cedida por um escritório de advocacia de Belo Horizonte/MG, que realiza cobrança extrajudicial (amigável) para

diversas instituições credoras em todo território nacional, seguindo as recomendações da LGPD para anonimização dos dados.

Trata-se de um portfólio de cartões de crédito Private Label (puro e híbrido) administrados por um dos maiores bancos do Brasil. De forma geral, cartões Private Label *puros* são linhas de crédito que podem ser utilizadas apenas na rede da instituição credora, dessa forma o cartão é emitido sem bandeira, com a identidade visual da empresa. Já os cartões Private Label *híbridos* são bandeirados, ou seja, podem ser utilizados na rede da instituição credora ou em qualquer outro lugar na função crédito.

Os inadimplentes em questão foram cedidos para a assessoria realizar a cobrança entre 01/01/2020 e 31/12/2020. Desconsiderou-se os clientes que negociaram suas dívidas e que a primeira parcela (ou parcela única) tinha vencimento a partir de janeiro de 2021, para não comprometer a separação dos grupos “bons” e “maus” pagadores. Além disso, clientes com dívidas menores que R\$10,00 (dez reais), também foram excluídos da base de dados, pois não são acionados pela equipe de cobrança da assessoria citada acima.

A base de dados é composta 665.379 (seiscentos e sessenta e cinco mil, trezentos e setenta e nove) clientes inadimplentes e por 23 (vinte e três) variáveis:

- Variável 1 - COD_CLIENTE: código para anonimização do inadimplente;
- Variável 2 - DT_ENTRADA: data em que o inadimplente foi incorporado a base de dados da assessoria de cobrança;
- Variável 3 - SEXO: sexo do inadimplente, podendo ser feminino ou masculino;
- Variável 4 - IDADE: idade do inadimplente;
- Variável 5 - ESTADO_CIVIL: estado civil do inadimplente, podendo ser casado(a) ou solteiro(a);
- Variável 6 - FASE: categorização utilizada pela instituição credora para classificar os clientes de acordo com o atraso da dívida.

FASE	ATRASO (DIAS)
FASE 02	91 A 150
FASE 03	151 A 180
FASE 04	181 A 360
FASE 05	361 A 720
FASE 06	721 A 1.080
FASE 07	1.081 A 1.440
FASE 08	1.441 A 1.800
FASE 09	1.801 A 2.160
FASE 10	> 2.160

- Variável 7 - ATRASO: faixa de atraso da dívida de acordo com o quadro acima;
- Variável 8 - SEGMENTO: tipo de cartão, podendo ser PRIVATE_LABEL (para cartões Private Label puro) ou PARCEIRO (para cartões Private Label híbrido);
- Variável 9 - VALOR_DIVIDA: valor atualizado da dívida do cliente, utilizado como referência durante a negociação;
- Variável 10 - FAIXA_DIVIDA: categorização utilizada pela assessoria de cobrança para classificar os clientes de acordo com o saldo da dívida.
 - Até 500,00
 - De 500,01 a 1.000,00
 - De 1.500,01 a 2.000,00
 - De 2.000,01 a 2.500,00
 - De 2.500,01 a 3.000,00
 - De 3.000,01 a 3.500,00
 - De 3.500,01 a 4.000,00
 - De 4.000,01 a 4.500,00
 - De 4.500,01 a 5.000,00
 - De 5.000,01 a 10.000,00
 - Maior que 10.000,00
- Variável 11 - DIA_VENC_CARTAO: dia de vencimento do cartão que originou a dívida;
- Variável 12 - STATUS_ACORDO: apontamento se houve negociação da dívida, podendo ser NEGOCIOU ou NÃO_NEGOCIOU;
- Variável 13 - ORIGEM_ACORDO: origem da emissão do acordo, podendo ser ASSESSORIA para dívidas negociadas pela equipe de cobrança da assessoria ou ESPONTANEO para acordos realizados diretamente nos canais de atendimento da instituição credora (banco);
- Variável 14 - VALOR_ACORDO: valor total do acordo realizado (somatório de todas as parcelas);
- Variável 15 - DESCONTO: percentual de desconto aplicado na negociação;
- Variável 16 - VALOR_PARCELA: valor da primeira ou única parcela do acordo;

- Variável 17 - PLANO: total de parcelas que compõe o acordo negociado, variando de 1 (um) a 25 (vinte e cinco) parcelas. A instituição credora permite negociar em, no máximo, entrada mais 24 (vinte e quatro) parcelas;
- Variável 18 - DT_EMISSAO: data de emissão do acordo, caso ocorra a negociação da dívida;
- Variável 19 - DT_VENCIMENTO: data de vencimento da primeira ou única parcela do acordo, caso ocorra a negociação da dívida;
- Variável 20 - PRAZO_PGMMTO1: prazo concedido para pagamento da primeira ou única parcela do acordo, caso ocorra a negociação da dívida;
- Variável 21 - DT_PAGAMENTO_ENTRADA: data do pagamento da primeira ou única parcela do acordo, caso não tenha ocorrido a quebra do acordo;
- Variável 22 - STATUS_PGMMTO: apontamento se houve pagamento da primeira ou única parcela do acordo, podendo ser PAGO ou QUEBRA;
- Variável 23 - CLASSIFICACAO: apontamento se o cliente é um bom ou mau pagador, podendo ser BOM_PAGADOR, caso tenha realizado o pagamento da primeira ou única parcela do acordo ou MAU_PAGADOR, caso o acordo tenha quebrado.

2.4 Método

A utilização da regressão logística para discriminação entre grupos relacionados ao risco de inadimplência é indicada por se tratar de uma variável resposta binária, podendo assumir somente dois valores $Y \in \{0, 1\}$, representando fracasso e sucesso.

De forma sucinta, pode-se representar a distribuição de probabilidade de Y como:

$$Y = \begin{cases} 1, & \text{com probabilidade } p. \\ 0, & \text{com probabilidade } (1 - p). \end{cases}$$

Para essa pesquisa, define-se como sucesso (valor de referência) o cliente bom pagador ($Y = 1$) e fracasso o cliente mau pagador ($Y = 0$).

A distribuição de probabilidade é definida como:

$$P(Y = y) = p^y(1 - p)^{1-y}$$

As chances de sucesso são definidas como a razão entre a probabilidade de sucesso e a probabilidade de fracasso.

$$\frac{p}{(1 - p)}$$

A interpretação do modelo é multiplicativa, a chance de um indivíduo com uma unidade a mais da variável preditora é e^β vezes a chance de um indivíduo com uma unidade a menos da variável preditora. Ou seja, a razão de chances é igual a e^β , sendo β igual ao coeficiente estimado.²

Uma ferramenta para avaliar a qualidade de predição de um modelo de regressão logística é a Curva ROC (*Receiver Operating Characteristics*)³. Para tal, faz-se necessário definir os conceitos de sensibilidade, especificidade e limiar.

- **Sensibilidade:** proporção de acertos das observações referentes a resposta $Y = 1$.
- **Especificidade:** proporção de acertos das observações referentes a resposta $Y = 0$.
- **Limiar:** ponto de corte para análise da sensibilidade. O limiar é diretamente proporcional a especificidade e inversamente proporcional a sensibilidade, podendo variar entre 0 e 1.

A Curva ROC representa o desenho da sensibilidade em função de $1 - \text{especificidade}$. A área do gráfico abaixo da Curva ROC é chamada de AUC e pode ser definida como uma medida do ajuste do modelo de regressão logístico. Geralmente,

² Descrição mais detalhada do método de regressão logístico e do conceito de razão de chances (OR – Odds ratio) pode ser obtida em Costa (2019).

³ Costa (2019) destina parte de um capítulo para abordar a análise da Curva ROC (*Receiver Operating Characteristics*).

os modelos com melhor ajuste apresentam valores maiores para a estatística AUC ($AUC \leq 1$).

De acordo com Barth (2004), utilizou-se a técnica *stepwise forward selection* para a escolha das variáveis que compuseram o modelo. Variáveis foram agregadas ou retiradas do modelo a cada passo, até chegar-se ao modelo final.

3 RESULTADOS E DISCUSSÕES

3.1 Caracterização da amostra

Dos 665.379 (seiscentos e sessenta e cinco mil, trezentos e setenta e nove) clientes inadimplentes que compõe a base de dados, 96,2% foram classificados como mau-pagadores e 3,8% como bons-pagadores. Isso demonstra um baixo índice de recuperação de crédito desse produto.

O público é predominantemente do sexo feminino (60%) e casado (72%), com idade média de 38 anos, mas variando entre 18 e 99 anos.

O valor das dívidas varia de R\$10,58 a R\$91.004,18, mas 75% delas está abaixo de R\$3.080,79 e possuem ticket mediano aproximado de R\$1.600,00. Além disso, 69% delas são oriundas de cartões Private Label puro.

A Fase 4 (181 a 360 dias de atraso) possui o maior volume de dívidas (24%), seguida pela Fase 6 (721 a 1.080 dias de atraso) com 14%. Conforme pode-se observar no Gráfico 1, a Fase 3 (151 a 180 dias de atraso) possui a menor variação entre os valores de dívidas, já a Fase 10 (> 2.160 dias de atraso) possui a maior variação.

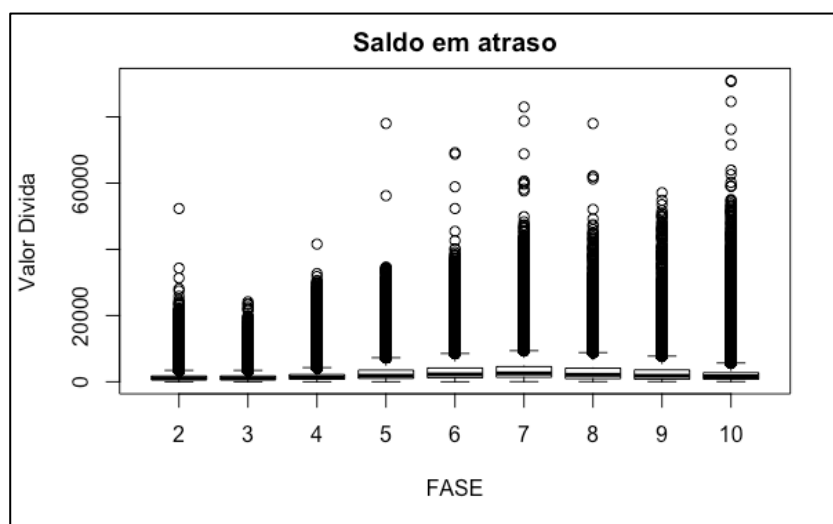


Gráfico 1: Valor das dívidas por fase (faixa de atraso)

Apenas 10% dos inadimplentes negociaram suas dívidas, sendo que 17,8% deles negociaram diretamente com os canais de atendimento do credor. A quebra foi de aproximadamente 62%, consideração apenas o pagamento da primeira parcela.

O ticket médio dos acordos é de aproximadamente R\$717,00, variando entre R\$10,40 e R\$24.015,40, mas 75% dos acordos são de até R\$1.250,00. As fases 4 e 2, respectivamente, possuem maior variação nos valores dos acordos, conforme pode-se verificar no Gráfico 2.

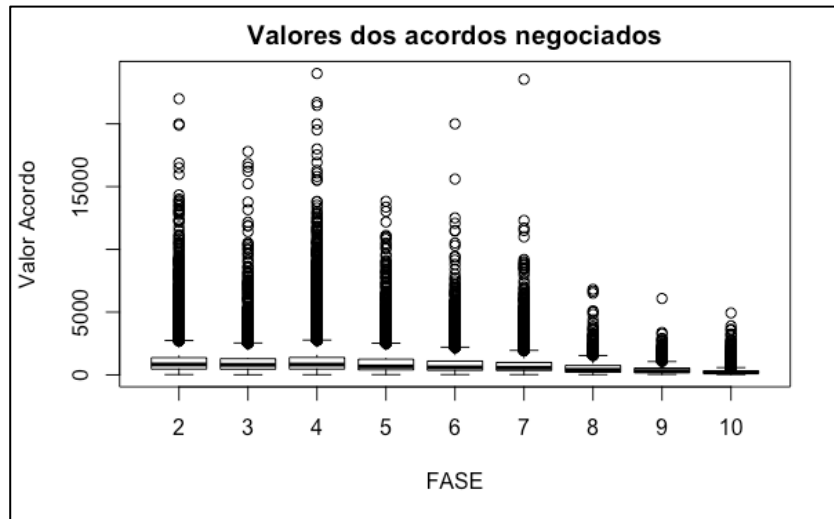


Gráfico 2: Valor dos acordos por fase (faixa de atraso)

O ticket médio das parcelas é de R\$371,00, sendo que 75% das parcelas possuem valor menor que R\$440,00. Através do Gráfico 3, percebe-se que as fases 9 e 10, respectivamente, possuem menor variação nos valores das parcelas.

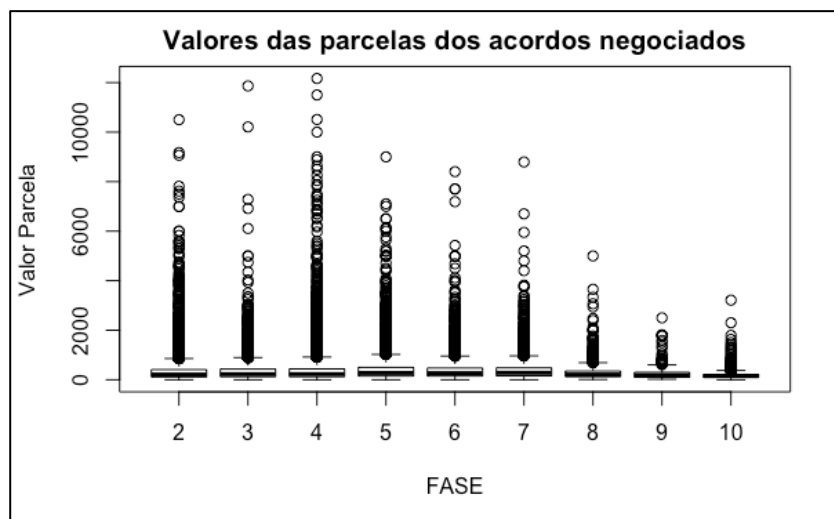


Gráfico 3: Valor da parcela por fase (faixa de atraso)

Dívidas mais velhas possuem maior margem para desconto, fato que pode ser verificado através do Gráfico 4, referente aos descontos aplicados.

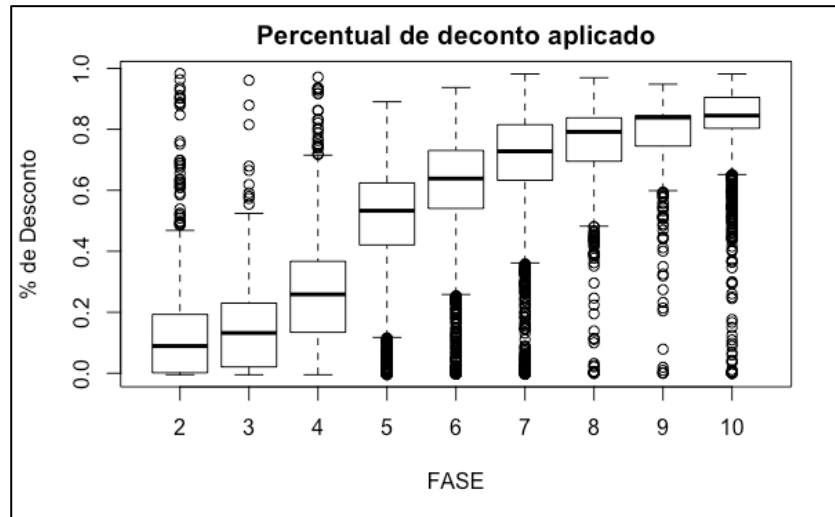


Gráfico 4: Desconto aplicado por fase (faixa de atraso)

Através do Gráfico 5, pode-se destacar alguns pontos relevantes:

- Acordos de maior valor possuem planos de parcelamento maiores;
- Acordos realizados em planos reduzidos tendem a ter maior desconto aplicado.

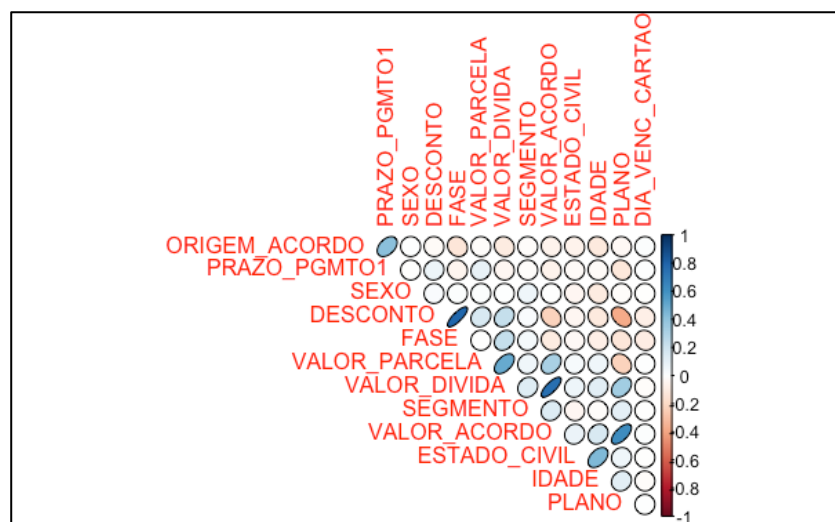


Gráfico 5: Análise de correlação

3.2 Ajuste do modelo de regressão logístico

O modelo de regressão logístico foi inicialmente ajustado com quatorze variáveis preditoras, sendo elas:

1. Variável 3 – SEXO (0 = Feminino e 1 = Masculino)
2. Variável 4 – IDADE
3. Variável 5 - ESTADO_CIVIL (0 = Casado e 1 = Solteiro)
4. Variável 6 – FASE
5. Variável 7 – ATRASO
6. Variável 8 – SEGMENTO (0 = Private Label e 1 = Parceiro)
7. Variável 9 – VALOR_DIVIDA
8. Variável 10 – FAIXA_DIVIDA
9. Variável 11 – DIA_VENC_CARTAO
10. Variável 13 – ORIGEM_ACORDO (0 = Assessoria e 1 = Espontâneo)
11. Variável 14 – VALOR_ACORDO
12. Variável 15 – DESCONTO
13. Variável 16 – VALOR_PARCELA
14. Variável 17 – PLANO e Variável 20 – PRAZO_PGMO1

Em seguida, foram excluídas todas as variáveis que possuem respostas em branco, pois elas desconsideram algumas observações para ajuste do modelo. Dessa forma, as variáveis relacionadas as informações sobre acordo foram desconsideradas, pois os clientes que não negociaram suas dívidas não tinham respostas válidas para esses campos.

Por fim, foram excluídas variáveis com p-valor maior que 0,05. O ajuste do modelo final considerou apenas as variáveis sexo, idade, estado civil, fase, valor da dívida e dia de vencimento do cartão, e seu resultado é apresentado a seguir:

```

modeloF <- glm(y ~ SEXO + IDADE + ESTADO_CIVIL + FASE + VALOR_DIVIDA + DIA_VENC_CARTAO,
family=binomial, data=dados)

Call:
glm(formula = y ~ SEXO + IDADE + ESTADO_CIVIL + FASE + VALOR_DIVIDA +
DIA_VENC_CARTAO, family = binomial, data = dados)

Deviance Residuals:

```



```

      Min       1Q   Median       3Q      Max
-0.5479 -0.3231 -0.2462 -0.1701  3.8630

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.394e+00  2.846e-02 -48.977 < 2e-16 ***
SEXO        -1.753e-01  1.345e-02 -13.037 < 2e-16 ***
IDADE       -1.794e-03  4.989e-04  -3.596 0.000323 ***
ESTADO_CIVIL  2.326e-01  1.572e-02  14.795 < 2e-16 ***
FASE        -3.004e-01  3.447e-03 -87.162 < 2e-16 ***
VALOR_DIVIDA -1.136e-04  3.819e-06 -29.752 < 2e-16 ***
DIA_VENC_CARTAO -3.967e-03  7.647e-04  -5.188 2.12e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 214974  on 665378  degrees of freedom
Residual deviance: 202596  on 665372  degrees of freedom
AIC: 202610

```

Com base nos resultados obtidos, segue o efeito de cada variável, dado que as outras estão fixas:

- **Efeito positivo:** somente a variável estado civil apresenta coeficiente positivo e contribui para o aumento da probabilidade de classificação como bom pagador. De acordo com a categoria de referência dessa variável (0 = Casado), podemos dizer que clientes casados tendem a ser bons pagadores.
- **Efeito negativo:** as variáveis com coeficientes negativos contribuem para a redução da probabilidade de classificação como bom pagador. Para a variável categórica sexo, de acordo com sua categoria de referência (0 = Feminino), podemos dizer que clientes do sexo feminino tendem a ser bons pagadores. Já para as variáveis numéricas (idade, fase, valor da dívida e dia de vencimento do cartão), podemos dizer que clientes com menores valores para essas variáveis tendem a ser bons pagadores.

As estimativas das razões de chance (OR) com os respectivos intervalos de confiança são apresentados abaixo:

```

exp(coef(modeloF))

(Intercept)          SEXO          IDADE          ESTADO_CIVIL          FASE
0.2480963          0.8392140          0.9982074          1.2618933          0.7405190
VALOR_DIVIDA DIA_VENC_CARTAO
0.9998864          0.9960405

exp(confint(modeloF))

```

	2.5 %	97.5 %
(Intercept)	0.2346312	0.2623241
SEXO	0.8173648	0.8616012
IDADE	0.9972299	0.9991821
ESTADO_CIVIL	1.2235593	1.3013428
FASE	0.7355228	0.7455273
VALOR_DIVIDA	0.9998789	0.9998938
DIA_VENC_CARTAO	0.9945489	0.9975346

Segue interpretação de cada variável numérica, mantendo-se as demais constantes:

- Idade: para cada aumento unitário nesta variável, diminui-se em 0,18% $((0,9982074 - 1) * 100)$ as chances de que o cliente seja um bom pagador.
- Fase: para a cada aumento unitário nesta variável, diminui-se em 25,9% $((0,740519 - 1) * 100)$ as chances de que o cliente seja um bom pagador.
- Valor da dívida: para a cada aumento unitário nesta variável, diminui-se em 0,01% $((0,9998864 - 1) * 100)$ as chances de que o cliente seja um bom pagador.
- Dia de vencimento do cartão: para a cada aumento unitário nesta variável, diminui-se em 0,40% $((0,9960405 - 1) * 100)$ as chances de que o cliente seja um bom pagador.

As curvas de sensibilidade e especificidade são mostradas na Figura 1 e a curva ROC é mostrada na Figura 2. A área abaixo da curva ROC é de 0.6996.

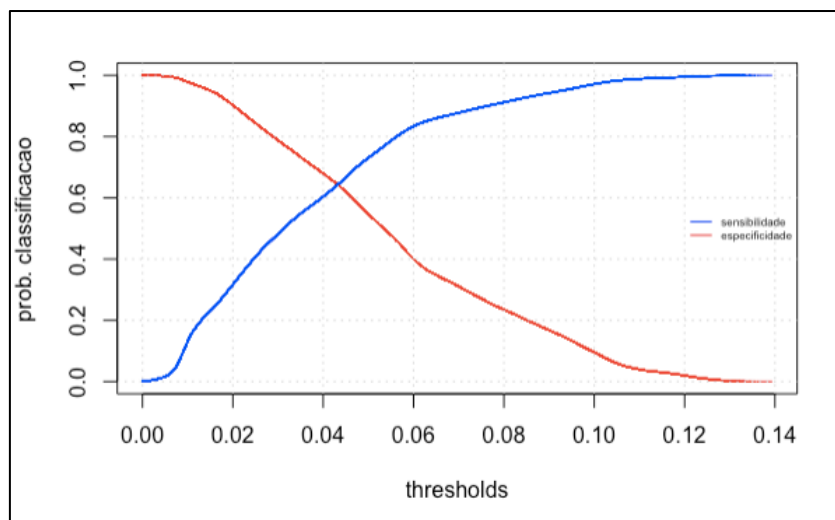


Figura 1: Curvas de sensibilidade e especificidade

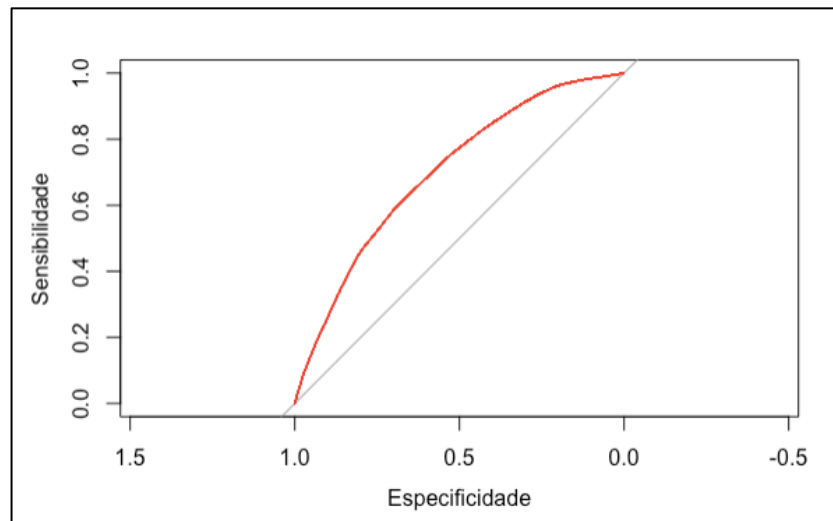


Figura 2: Curva ROC

Através do código abaixo, estimou-se a probabilidade de cada cliente inadimplente da base tornar-se adimplente, possibilitando a criação de um ranking de propensão de negociação.

```
require(WriteXLS)

dados$Prob <- predict(modeloF, type="response")

final <- dados[,c("SEXO", "IDADE", "ESTADO_CIVIL", "FASE",
                 "VALOR_DIVIDA",
                 "DIA_VENC_CARTAO", "Prob")]
final <- unique(final)
WriteXLS(final, "Saida.xlsx")
```

A observação com maior propensão de se tornar um bom pagador possui 13,94% de probabilidade de negociar sua dívida e pagar a primeira parcela do acordo. Refere-se a uma pessoa do sexo feminino, com 22 anos de idade, solteira, com uma dívida de R\$132,79 da Fase 2 (91 a 150 dias de atraso) e com cartão vencendo dia 01.

Já a observação com menor propensão de se tornar um bom pagador possui 0,00003% de probabilidade de negociar sua dívida e pagar a primeira parcela do acordo. Também se refere a uma pessoa do sexo feminino, mas com 88 anos de idade, casada, com uma dívida de R\$91.004,18 (a maior dívida da base de dados) da Fase 4 (181 a 360 dias de atraso) e com cartão vencendo dia 04.

4 CONCLUSÃO

Os resultados mostram que o modelo logístico ajustado, apesar de não apresentar grande capacidade de discriminação entre os grupos de clientes *bons pagadores* e *maus pagadores*, reflete a realidade do negócio e a base de dados apresentada, ou seja, para se acertar um bom pagador é necessário tentar negociar com muitos maus pagadores.

Apesar disso, trata-se de uma ferramenta que pode agregar na construção de réguas de cobrança mais eficientes, através da priorização de clientes com maior probabilidade de efetivarem negociações.

Além de melhorar o gerenciamento do negócio, a utilização de modelos de score permite aumento da efetividade financeira, gerando redução dos custos com cobranças improdutivas.

Vale ressaltar que as bases de cobrança mudam constantemente e podem ser impactadas por fatores econômicos, políticos, operacionais, entre outros. Portanto, os modelos de score sofrem influência das alterações de perfil do público-alvo.

Outro ponto pertinente está relacionado ao conjunto restrito de informações que compõe a base de dados para análise. Deve-se considerar que outras variáveis podem influenciar a condição do cliente inadimplente se tornar adimplente, tais como renda, se está empregado ou não, etc. Esse tipo de informação pode auxiliar no ajuste do modelo, mas nem sempre estarão disponíveis, seja em função do custo para obtê-las ou pela disponibilidade das mesmas.

Dessa forma, o desafio é monitorar e calibrar sempre que necessário o modelo, para obter resultados mais assertivos e com maior impacto ao negócio.

REFERÊNCIAS

- BANCO CENTRAL DO BRASIL. Resolução nº 2682, de 21 de dezembro de 1999. Dispõe sobre critérios de classificação das operações de crédito e regras para constituição de provisão para créditos de liquidação duvidosa. Disponível em: https://www.bcb.gov.br/pre/normativos/res/1999/pdf/res_2682_v2_L.pdf. Acesso em: 21 jan. 2021.
- BANCO CENTRAL DO BRASIL. Recomendações de Basileia. Disponível em: <https://www.bcb.gov.br/estabilidadefinanceira/recomendacoesbasileia>. Acesso em: 21 jan. 2021.
- BARTH, Nelson Lerner. Inadimplência: construção de modelos de previsão. São Paulo: Nobel, 2004.
- BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/ato2015-2018/2018/lei/L13709.htm. Acesso em: 21 jan. 2021.
- COSTA, Marcelo Azevedo. **Tópicos em ciências de dados: introdução dos modelos paramétricos e suas aplicações utilizando o R**. Rio de Janeiro: Bonecker, 2019.
- DELOITTE. Maturidade da Competência de Cobrança. 2011. Disponível em: http://www.portalabbc.org.br/arquivos/2011_05_02_collection_score_abbc.pdf. Acesso em: 21 jan. 2021.
- MACHADO, Aline Rodrigues. **Collection Scoring via Regressão Logística e Modelo de Riscos Proporcionalis de Cox**. 2015. 98 f. Dissertação (Mestrado em Estatística) – Departamento de Estatística do Instituto de Ciências Exatas, Universidade de Brasília, Brasília, 2015.
- MARQUES, Rosana. **Cobrança e a LGPD: como recuperar crédito sem infringir a Lei?** Disponível em: <https://blog.assertivasolucoes.com.br/cobranca-e-a-lgpd/>. Acesso em: 21 jan. 2021.
- SILVA, Eli Lopes da. **Elaboração de trabalhos acadêmicos [livro eletrônico]: normas, dicas e erros comuns** / Eli Lopes da Silva. – 3. ed. – Florianópolis: Terçária Tecnologias Educacionais, 2021. 131 p. : il. ; PDF ; 3,48 MB.
- Métodos Estatísticos para Análise de Dados e Recuperação Estratégica**. Disponível em: <https://collbusinessnews.com.br/metodos-estatisticos-para-analise-de-dados-e-recuperacao-estrategica/>. Acesso em: 17 dez. 2020.
- SERASA EXPERIAN. **Mapa de inadimplência do 1o. semestre de 2020: uma análise dos impactos da pandemia nas ações de cobrança**. [2020]. *E-book*. Disponível em: <https://www.serasaexperian.com.br/conteudos/cobranca/mapa-da-inadimplencia-do-1o-semester-de-2020-uma-analise-dos-impactos-da-pandemia-nas-acoes-de-cobranca/>. Acesso em: 21 jan. 2021.