CLASSIFICAÇÃO DE CLIENTES INADIMPLENTES E GERAÇÃO DE INFORMAÇÃO PARA A TOMADA DE DECISÃO EM CRÉDITO: UMA ANÁLISE SOB A ÓTICA DA CIÊNCIA DA INFORMAÇÃO

Classification of Delinquent Customers and Information Generation for Credit Decision-Making: An Analysis from the Perspective of Information Science

JULIANA SIMÕES DA SILVA DUTRA
UNIVERSIDADE FEDERAL DE MINAS GERAIS (UFMG)

FREDERICO CESAR MAFRA PEREIRA
UFMG

MARCELO AGENOR ESPÍNDOLA
FACULDADE DE NOVA SERRANA (FANS)

Comunicação:
O XIV SINGEP foi realizado em conjunto com a 14th Conferência Internacional do CIK (CYRUS Institute of Knowledge) e com a SPARK (Society for Promoting Academic Research and Knowledge), em formato híbrido, com sede presencial na UNINOVE - Universidade Nove de Julho, no Brasil.

CLASSIFICAÇÃO DE CLIENTES INADIMPLENTES E GERAÇÃO DE INFORMAÇÃO PARA A TOMADA DE DECISÃO EM CRÉDITO: UMA ANÁLISE SOB A ÓTICA DA CIÊNCIA DA INFORMAÇÃO

Objetivo do estudo
O presente trabalho tem como objetivo identificar variáveis relacionadas à probabilidade de adimplência de clientes inadimplentes, além de discutir o papel da gestão da informação e inteligência analítica na construção de estratégias de cobrança mais eficientes.

Relevância/originalidade
O trabalho é relevante e original ao analisar a inadimplência sob a perspectiva da Ciência da Informação, empregando técnicas estatísticas para desenvolver um modelo preditivo que apoia estratégias de recuperação de crédito e tomada de decisão.

Metodologia/abordagem
A pesquisa caracteriza-se como quantitativa, descritiva e exploratória. A base de dados, anonimizada e em conformidade com a LGPD, compreendeu informações de 665.379 clientes inadimplentes de cartões de crédito.

Principais resultados
Os resultados demonstraram que 96,2% dos clientes foram classificados como maus pagadores e 3,8% como bons pagadores. Estado civil, fase da dívida, valor da dívida, idade, sexo e dia de vencimento foram as variáveis mais relevantes. Modelo apresentou AUC de 0,70.

Contribuições teóricas/metodológicas
Este trabalho contribui para com as metodologias de desenvolvimento de modelos preditivos sob a perspectiva da Ciência da Informação, aplicado à recuperação de crédito.

Contribuições sociais/para a gestão
Este trabalho mostrou que a integração entre Ciência da Informação, Estatística e Ciência de Dados favorece estratégias de cobrança mais eficientes, segmentadas e alinhadas à gestão do conhecimento organizacional.

Palavras-chave: Comportamento do Consumidor, Informação para Tomada de Decisão, Inadimplência, Gestão da Informação, Recuperação de crédito

Classification of Delinquent Customers and Information Generation for Credit Decision-Making: An Analysis from the Perspective of Information Science

Study purpose
This study aims to identify variables related to the likelihood of delinquent customers settling their debts, as well as to discuss the role of information management and analytical intelligence in developing more efficient collection strategies.

Relevance / originality
The study is relevant and original in its analysis of loan default from the perspective of Information Science, employing statistical techniques to develop a predictive model that supports credit recovery strategies and decision-making.

Methodology / approach
The research is characterized as quantitative, descriptive, and exploratory. The dataset—anonymized and compliant with the LGPD—comprised information on 665,379 delinquent credit card customers.

Main results
The results showed that 96.2% of customers were classified as bad payers and 3.8% as good payers. Marital status, debt stage, debt amount, age, gender, and due date were the most relevant variables. The model yielded an AUC of 0.70.

Theoretical / methodological contributions
This study contributes to predictive model development methodologies from the perspective of Information Science, applied to credit recovery.

Social / management contributions
This study demonstrated that the integration of Information Science, Statistics, and Data Science fosters collection strategies that are more efficient, segmented, and aligned with organizational knowledge management.

Keywords: Consumer Behavior, Information for Decision-Making, Default, Information Management, Debt recovery

CLASSIFICAÇÃO DE CLIENTES INADIMPLENTES E GERAÇÃO DE INFORMAÇÃO PARA A TOMADA DE DECISÃO EM CRÉDITO: UMA ANÁLISE SOB A ÓTICA DA CIÊNCIA DA INFORMAÇÃO

1 Introdução

A inadimplência é um dos principais desafios econômicos do Brasil atual. Segundo dados do Serasa Experian, em setembro de 2025, o país registrou 79,1 milhões de endividados, representando 42,4% da população adulta. Já entre as empresas brasileiras, a inadimplência bateu recorde em agosto de 2025, chegando a 8,1 milhões de CNPJs negativados. Esses dados mostram a magnitude do problema e seus impactos sistêmicos sobre o crédito, o consumo e a sustentabilidade financeira das organizações.

Diante desse cenário, atuar de forma assertiva e ágil com a recuperação de crédito é um diferencial competitivo em empresas para melhorar sua saúde financeira. Compreender o perfil dos devedores e desenvolver estratégias baseadas em dados é essencial para a eficiência das instituições financeiras e para a formulação de políticas públicas de educação financeira.

O objetivo deste artigo é utilizar análise de dados e aplicar um modelo estatístico de regressão logística para classificar os clientes inadimplentes de uma base de cartões de crédito Private Label de acordo com a propensão de negociação de sua dívida, possibilitando melhorar as estratégias de cobrança, além de discutir sua contribuição para a Ciência da Informação e para a gestão organizacional.

O estudo justifica-se pela necessidade crescente de estudos e análises preditoras capazes de avaliar os riscos na concessão de empréstimos e na recuperação de dívidas, já que 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.

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.

Dentre os principais modelos preditivos relacionados ao processo de cobrança pode-se citar o Collection Scoring[i], 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.

Sob a ótica da Ciência da Informação (CI), a inadimplência pode ser entendida como um fenômeno informacional: envolve tratamento de dados e geração de informações para tomada de decisão. Dessa forma, a análise de dados e a utilização de modelos estatísticos tornam-se instrumentos fundamentais para gerar conhecimento e orientar estratégias de cobrança, concessão e renegociação de dívidas.

2 Referencial Teórico

De acordo com Davenport (1998), a CI insere-se em um contexto histórico em que a informação e o conhecimento se tornam elementos centrais da sociedade e da economia. Ele destaca que vivemos em uma sociedade na qual a informação é tratada como uma "commodity valiosa" e um "grande negócio", de modo que a informação de qualidade passa a ser um recurso estratégico, que deve ser gerido com rigor, integridade e intencionalidade dentro das organizações. Para o autor, a CI é compreendida como um campo de investigação científica e prática profissional que se desenvolveu ao longo do último meio século, marcada por três características gerais: (i) **Natureza interdisciplinar:** a CI se conecta com várias disciplinas, como por exemplo administração, ciência da computação, sociologia, etc.; (ii) **Tecnologia da informação:** a CI está conectada a tecnologia, que pode impulsionar ou restringir a forma como a informação é capturada, armazenada, processada e distribuída; (iii) **Participação na chamada sociedade da informação:** a CI possui forte dimensão social e humana, enfatizando usos, contextos, culturas organizacionais e práticas de trabalho.

Davenport (1998) contrapõe duas abordagens que, historicamente, disputaram espaço na área: a abordagem centrada no sistema, focada em algoritmos e modelos tradicionais de recuperação da informação e a abordagem centrada no usuário, que enfatiza aspectos cognitivos, interativos e contextuais. A integração bem-sucedida dessas abordagens é vista como caminho para o progresso da CI. Um ponto central na ecologia da informação proposta por Davenport (1998) é a distinção entre dados, informação e conhecimento. Os dados são descritos como "observações sobre o estado do mundo", facilmente estruturados, quantificados, processados por máquinas e transferíveis, sem a necessidade de intervenção humana. A informação é formada por "dados dotados de relevância e propósito", o que exige análise, escolha de formas de representação e intervenção humana. Mesmo quando uma máquina gera gráficos ou dashboards, há sempre a necessidade de decisão humana, escolhendo quais dados utilizar, quais indicadores destacar e qual narrativa construir. Já o conhecimento representa a forma mais valiosa de informação, pois incorpora reflexão, síntese, experiência e contexto, sendo difícil de capturar integralmente em sistemas automatizados. De forma resumida, a análise é o processo que media a passagem de dados para informação, dotando-os de significado e utilidade para a tomada de decisão (Davenport, 1998).

A análise é um passo importante no processo de gerenciamento da informação que Davenport (1998) descreve como um conjunto estruturado de atividades, no qual a informação percorre um ciclo estruturado desde sua identificação, até seu uso efetivo para a tomada de decisão. O modelo genérico desse processo é formado por quatro etapas: (i) determinação de exigências, (ii) coleta, (iii) distribuição e (iv) utilização, mas Davenport (1998) também cita o modelo de oito etapas, utilizado pela empresa britânica Standard Life Assurance[ii]: (i) identificação, (ii) coleta, (iii) organização, (iv) análise, (v) interpretação, (vi) distribuição e (vii) utilização. Davenport (1998) critica a análise puramente tecnológica, que produz informações com pouco contexto ou significado. Computadores são úteis para lidar com dados, mas a atuação humana é essencial para contextualizar, interpretar e comunicar resultados de maneira aplicável à solução de problemas e ao apoio a tomada de decisões. Dessa forma, a análise de dados deve ser vista como um processo humano que dá significado, contexto e relevância aos dados, para que eles sejam utilizados de forma eficaz nos processos decisórios das organizações.

Já para o autor Choo (2006), a CI dedica-se a compreender os processos de geração, organização, mediação e uso da informação para apoiar decisões mais racionais em contextos de incerteza. O autor reforça que a informação conecta o conhecimento às ações organizacionais. Em sua proposta de "organização do conhecimento", o autor destaca três arenas centrais: (i) **Criação de significados (sensemaking):** arena em que a organização

interpreta as informações sobre o ambiente, para compreender mudanças, ameaças e oportunidades; (ii) **Construção do conhecimento (knowledge creation):** nesta arena, a organização cria, organiza e processa a informação para gerar novos conhecimentos, a partir de aprendizagem e inovação; (iii) **Tomada de decisão (decision making):** por fim, nesta arena, a organização processa e analisa a informação para tomada de decisão.

No contexto da inadimplência, a arena da criação de significados (sensemaking) envolve o monitoramento contínuo de informações econômicas, financeiras, comportamentais e institucionais, como índices de endividamento, variações na renda da população, alterações no cenário macroeconômico e mudanças regulatórias. Através do sensemaking, esses dados são interpretados e colaboram com explicações sobre as causas, tendências e padrões do não pagamento. Dessa forma, pode-se compreender por que determinados perfis se tornam mais propensos à inadimplência. A arena da construção do conhecimento (*knowledge creation*) materializa-se, no contexto da inadimplência, por meio da necessidade que as organizações têm de integrar diversas bases de dados de forma estruturada, para processá-las através de ferramentas analíticas e modelos estatísticos. Por fim, no contexto da inadimplência, a arena da tomada de decisão (*decision making*) é o momento em que as organizações definem as estratégias relacionadas à concessão de crédito, à definição de limites, à segmentação de carteiras, à priorização de clientes para cobrança, às condições de renegociação e à alocação de recursos operacionais. Contudo, conforme destaca Choo (2006), os modelos estatísticos subsidiam e qualificam o julgamento humano, mas não o substituem. Dessa forma, a integração entre as três arenas forma um ciclo contínuo no qual as organizações recuperadoras de crédito podem interpretar o fenômeno da inadimplência, através da transformação de dados em conhecimento por meio da regressão logística e da utilização desse conhecimento para fundamentar decisões mais assertivas na gestão do risco de crédito.

Sob a ótica da CI, a regressão logística deve ser compreendida como parte integrante de um ecossistema informacional composto por bases de dados, sistemas de informação, processos organizacionais e práticas de uso da informação (Davenport, 1998; Choo, 2006). O valor informacional do modelo não reside apenas nos coeficientes matemáticos estimados, mas na capacidade de integrar seus resultados aos fluxos decisórios da organização, por meio de relatórios, dashboards, sistemas de cobrança e políticas de crédito. Bawden e Robinson (2012) contribuem para esse quadro ao definirem a CI como a ciência e a prática relacionadas à coleta, armazenamento, recuperação e uso eficaz da informação registrada e do conhecimento. Para esses autores, a informação é um conceito complexo e multifacetado, cuja compreensão exige dialogar com diferentes tradições, incluindo a teoria matemática da comunicação de Shannon e a filosofia da informação de Floridi.

A Teoria Matemática da Comunicação, formulada por Claude Shannon em 1948, representa o marco da formalização quantitativa do conceito de informação. Shannon mostrou que toda comunicação pode ser vista como um processo de envio de sinais, que pode ser medido, calculado e aprimorado através da matemática. O autor relacionou informação ao conceito de entropia. Nessa perspectiva, quanto maior a imprevisibilidade de um evento, maior é a quantidade de informação gerada quando ele ocorre. Embora sua teoria se concentre nos níveis empírico e sintático da informação, sem considerar seu significado ou valor social, ela forneceu o suporte estatístico e tecnológico para o desenvolvimento dos sistemas de comunicação e dos atuais sistemas de informação digitais (Shannon, 1948).

No campo da CI, a teoria de Shannon é reconhecida como fundamento técnico para o tratamento, a transmissão e o armazenamento da informação, ao oferecer os princípios matemáticos que sustentam a codificação, a compressão e a circulação de dados em ambientes digitais (Bawden; Robinson, 2012). No entanto, a CI amplia essa abordagem, pois considera a interpretação, o contexto de uso e a apropriação social e organizacional da informação. Dessa forma, enquanto Shannon estrutura a base quantitativa dos fluxos informacionais, a CI

compreende como esses fluxos são organizados, mediados e utilizados para a produção de conhecimento e para a tomada de decisão nas organizações.

No contexto da análise de crédito e da inadimplência, essa integração torna-se ainda mais evidente. Empresas de cobrança operam em ambientes marcados por alto grau de incerteza quanto ao comportamento futuro dos clientes. Os modelos de regressão logística permitem estimar probabilidades de inadimplência com base em dados históricos e comportamentais, reduzindo a entropia do sistema decisório e qualificando a alocação de recursos, a segmentação de carteiras e a definição de estratégias de recuperação. A Ciência da Informação, por sua vez, assegura que esses dados sejam coletados, organizados, analisados e usados de forma estruturada e orientada ao conhecimento, potencializando o impacto do modelo estatístico na tomada de decisão estratégica.

Já a Filosofia da Informação, proposta por Luciano Floridi, define a informação como dados bem formados, significativos e verdadeiros, dotados de valor semântico e epistêmico (Floridi, 2011). Essa concepção amplia significativamente o escopo da CI, ao situar a informação como fundamento ontológico da realidade social e digital, estruturando aquilo que o autor denomina infosfera: o ambiente informacional no qual os seres humanos e os sistemas artificiais coexistem e interagem.

No âmbito da CI, a abordagem de Floridi aprofunda a dimensão semântica e ética da informação, complementando as visões técnicas e organizacionais propostas por autores como Shannon, Davenport e Choo. Enquanto a Teoria Matemática da Comunicação estrutura a informação em termos de transmissão e redução de ruído, e os modelos de gestão da informação enfatizam a organização e o uso estratégico dos dados, a Filosofia da Informação aborda a qualidade, o significado, a verdade e a responsabilidade no tratamento da informação (Floridi, 2011). Assim, a CI passa a ser compreendida não apenas como ciência dos sistemas, mas como ciência da significação, da mediação e do impacto social da informação. Essa perspectiva torna-se relevante quando analisamos o uso de modelos estatísticos no contexto organizacional. Sob a ótica da Filosofia da Informação, a regressão logística pode ser compreendida como uma ferramenta de transformação de dados em informação semanticamente qualificada, desde que seja alimentado com dados que atendam a critérios de boa formação, significado e verdade. Se um modelo estatístico for alimentado por dados incompletos, enviesados ou descontextualizados, comprometerá a qualidade do conhecimento gerado e das decisões tomadas. Essas decisões também precisam ser fundamentadas na responsabilidade informacional, ou seja, devem proteger a dignidade informacional dos indivíduos e garantir o uso justo, transparente e responsável dos dados que compõem a infosfera (Floridi, 2011).

Nesse contexto, a CI atua como mediadora entre os fundamentos filosóficos da informação e a operacionalização técnica dos modelos analíticos. Ao organizar, classificar, armazenar, recuperar e disponibilizar dados e informações, a CI garante que os insumos utilizados na regressão logística atendam aos critérios de qualidade informacional, possibilitando que os resultados do modelo sejam não apenas estatisticamente válidos, mas também semanticamente significativos e eticamente aceitáveis (Bawden; Robinson, 2012; Floridi, 2011). Portanto, a integração entre sistemas de informação, processos de gestão de dados e práticas analíticas passa a ser compreendida como parte da curadoria da infosfera organizacional.

Para Bawden e Robinson (2012), a organização da informação e a tecnologia são componentes estruturantes da CI. A organização da informação permite que grandes volumes de dados e documentos sejam descritos, relacionados e recuperados, enquanto a tecnologia viabiliza a coleta, o gerenciamento, a análise e a visualização desses dados, ampliando a capacidade das organizações de produzir conhecimento e apoiar decisões.

Sob a ótica da ecologia da informação, a implantação de um modelo de regressão logística em uma empresa de cobrança implica desenhar processos que garantam a qualidade

dos dados de entrada, a atualização contínua dos parâmetros do modelo, a integração dos resultados em sistemas operacionais e o retorno sobre o uso das informações pelos gestores e operadores de cobrança (Davenport, 1998).

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 CMN no 4.966, de 25 de novembro de 2021 do Banco Central do Brasil, que "dispõe sobre classificação, mensuração, perdas esperadas e demais requisitos relativos a instrumentos financeiros pelas instituições autorizadas pelo Banco Central do Brasil" determina critérios que devem ser utilizados nesse processo.

> A avaliação da perda esperada é de responsabilidade da instituição detentora do instrumento e deve ser efetuada com base em critérios consistentes e passíveis de verificação, amparada por informações internas e externas. (Banco Central do Brasil, 2021).

A resolução moderniza a gestão do risco de crédito ao exigir que as instituições financeiras adotem uma abordagem preditiva para mensurar perdas esperadas, que representam estimativas do valor que a instituição tende a não receber ao longo da vida dos instrumentos financeiros, considerando o risco de deterioração futura, mesmo que não exista atraso atual. Esse processo exige tratamento de dados, análise de séries históricas, integração entre sistemas, curadoria informacional, governança e qualidade da informação. Nessa perspectiva, a resolução aproxima o setor financeiro das práticas de gestão da informação, nas quais o ciclo informacional (coleta, organização, análise, armazenamento e uso) é essencial para transformar dados brutos em conhecimento acionável, capaz de sustentar decisões de negócio.

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.

Em síntese, a análise de dados e o uso de modelos estatísticos devem ser vistos como ferramentas centrais de uma estratégia mais ampla de gestão informacional. Em um ambiente

de inadimplência elevada, essa integração entre organização da informação, tecnologia, Estatística e processos decisórios permite que organizações transformem grandes volumes de dados em conhecimento aplicado, reduzam a incerteza, qualifiquem o julgamento humano e tomem decisões mais assertivas, estratégicas e alinhadas à racionalidade intencional que a CI busca promover.

3 Metodologia

A pesquisa adota uma abordagem quantitativa, de natureza descritiva e exploratória (Gil, 2008). 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. Por questões de confidencialidade, o escritório não será identificado.

O método estatístico aplicado foi a regressão logística[iii] (Cox, 1970), que permite estimar a probabilidade de um cliente inadimplente tornar-se adimplente (bom pagador). As variáveis independentes foram selecionadas por meio da técnica *stepwise forward selection*[iv] (Efroymson, 1960) e o modelo foi ajustado no Software R. Para avaliação do desempenho, utilizou-se a Curva ROC - *Receiver Operating Characteristic*[v] (Swets, 1988), considerando a área sob a curva (AUC) como indicador de qualidade preditiva.

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.

- **Bom pagador:** negociou sua dívida e realizou o pagamento da primeira parcela;
- **Mau pagador:** não negociou sua dívida ou negociou, mas não efetivou o pagamento da primeira parcela (quebra do acordo).

O estudo utiliza dados de uma carteira de 665.379 (seiscentos e sessenta e cinco mil, trezentos e setenta e nove) clientes inadimplentes de cartões de crédito Private Label (puros e híbridos) administrados por um dos maiores bancos do Brasil, com dívidas acima de R$10,00 (dez reais). As variáveis consideradas abrangem informações demográficas, financeiras e comportamentais, tais como sexo, idade, estado civil, fase da dívida, valor da dívida, dia de vencimento do cartão, entre outras. 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.

A base de dados é composta 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 02 – de 91 a 150 dias
    - Fase 03 – de 151 a 180 dias
    - Fase 04 – de 181 a 360 dias
    - Fase 05 – de 361 a 720 dias

  - Fase 06 – de 721 a 1.080 dias
  - Fase 07 – de 1.081 a 1.440 dias
  - Fase 08 – de 1.441 a 1.800 dias
  - Fase 09 – de 1.801 a 2.160 dias
  - Fase 10 – acima de 2.160 dias
- 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_PGMTO1: 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_PGMTO: apontamento se houve pagamento da primeira ou única parcela do acordo, podendo ser PAGO ou QUEBRA;

- Variável 23 - CLASSIFICAÇAO: 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.

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:

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

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.[vi]

Uma ferramenta para a avaliar a qualidade de predição de um modelo de regressão logística é a Curva ROC (*Receiver Operating Characteristics*). 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-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, os modelos com melhor ajuste apresentam valores maiores para a estatística AUC (AUC $\leqslant$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.

4 Análise dos resultados e Discussões

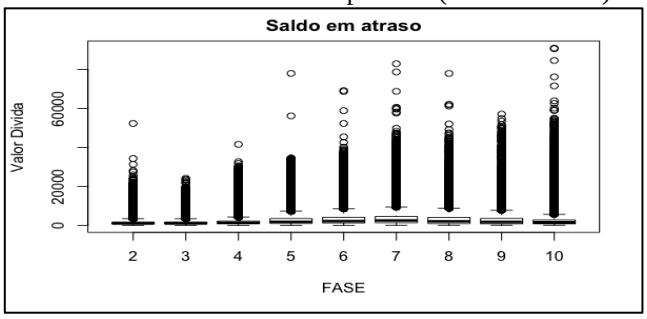
Os resultados indicam que 96,2% dos clientes analisados foram classificados como maus pagadores e apenas 3,8% como bons pagadores, demonstrando um baixo índice de recuperação de crédito desse produto. Outras análises podem ser feitas posteriormente para avaliar se a baixa recuperação é uma característica desse produto ou se estratégia de cobrança realizada não está adequada para esse público. A predominância de dívidas com mais de 180 dias de atraso sugere a necessidade de estratégias de cobrança mais segmentadas e baseadas em dados. As variáveis com maior influência na probabilidade de adimplência foram estado civil, fase da dívida e valor da dívida.

O modelo logístico apresentou uma área sob a curva ROC de 0,70, o que indica discriminação moderada. Os clientes casados e com dívidas de menor valor apresentaram maior

probabilidade de pagamento. Em termos informacionais, esses resultados mostram o potencial da análise de dados para segmentar perfis e priorizar esforços de recuperação de crédito, otimizando os custos operacionais. Além disso, o desenvolvimento de uma plataforma de autosserviço podem ser uma oportunidade para direcionar o público com alta propensão de negociação, gerando mais eficiência operacional. 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.

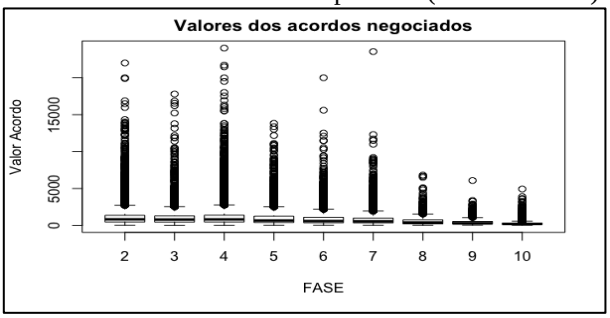
Gráfico 1- Valor das dívidas por fase (faixa de atraso)



Fonte: Dados da pesquisa (2026)

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. Considerando o alto índice de quebra de acordo, sugere-se que avaliações qualitativas sejam realizadas pela empresa, através de monitorias de ligações e atendimentos digitais, a fim de entender a consistência das negociações realizadas. 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.

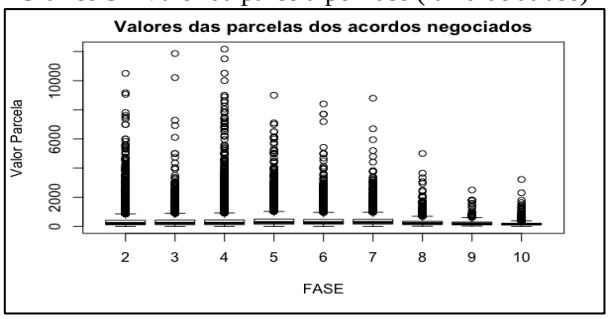
Gráfico 2 - Valor dos acordos por fase (faixa de atraso)



Fonte: Dados da pesquisa (2026)

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.

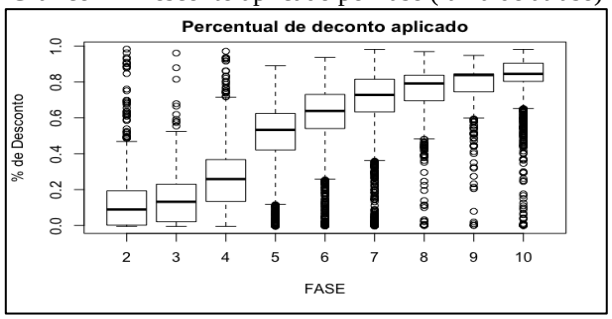
Gráfico 3 - Valor da parcela por fase (faixa de atraso)



Fonte: Dados da pesquisa (2026)

Dívidas mais velhas possuem maior margem para desconto, fato que pode ser verificado através do Gráfico 4, referente aos descontos aplicados. Utilizar essa informação pode agregar na argumentação durante a negociação com o cliente, reforçando a vantagem de realizar o acordo de pagamento.

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

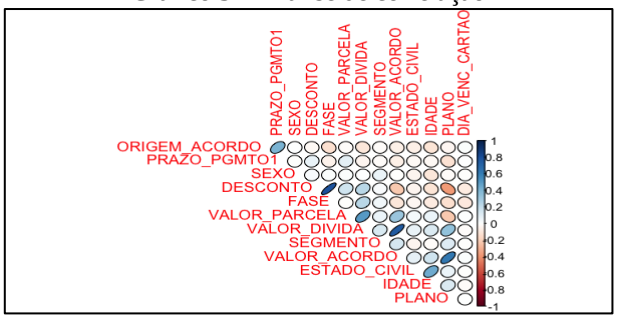


Fonte: Dados da pesquisa (2026)

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



Fonte: Dados da pesquisa (2026)

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_PGMTO1

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:

Figura 1 – Modelagem Ajustada

```
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
```

Fonte: Dados da pesquisa (2026)

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.

Essas informações devem ser utilizadas para direcionar a criação das estratégias de cobrança, como por exemplo, a priorização dos clientes com perfil de bom pagador e o direcionamento dos clientes com perfil de mau pagador para uma célula de atendimento especializada em quebra de objeções ou com perfil mais sênior.

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

Figura 2 – Estimativa e intervalos de confiança

```
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
```

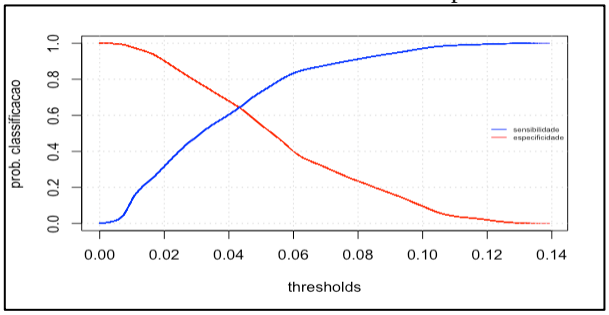
Fonte: Dados da pesquisa (2026)

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.

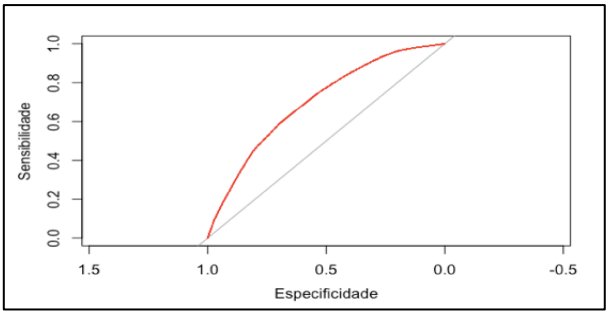
As curvas de sensibilidade e especificidade são mostradas nos Gráficos 6 e a curva ROC é mostrada no Gráfico 7. A área abaixo da curva ROC é de 0.6996.

Gráfico 6 - Curvas de sensibilidade e especificidade



Fonte: Dados da pesquisa (2026)

Gráfico 7 - Curva ROC



Fonte: Dados da pesquisa (2026)

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.

Figura 3 – Modelagem para 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")
```

Fonte: Dados da pesquisa (2026)

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.

Em síntese, a análise descritiva dos dados, sob a perspectiva da CI, possibilitou a organização, a representação e a interpretação sistemática do perfil da base de cobrança, convertendo dados dispersos em informação estruturada e socialmente significativa para a organização. Esse processo permitiu à empresa estabelecer relações comparativas com fontes externas de referência, como o Mapa da inadimplência e renegociação de dívidas do Serasa Experian[vii], bem como realizar práticas de benchmarking informacional com outras instituições do setor. Além disso, a estruturação dessa base informacional resultou na consolidação de um repositório de conhecimento organizacional, que pode ser utilizado no treinamento de agentes de atendimento baseados em inteligência artificial generativa, fortalecendo os processos de mediação da informação, aprendizagem organizacional e padronização do atendimento.

Já o ranking de clientes produzido a partir do modelo de regressão logística pode ser interpretado como um produto informacional resultante da transformação de dados em conhecimento aplicado ao processo decisório. A informação gerada pelo modelo subsidia a construção de estratégias de cobrança mais personalizadas, orientadas por evidências e alinhadas aos princípios da gestão da informação e do conhecimento. A partir dessa segmentação informacional, tornam-se viáveis ações como a implementação de plataformas de autosserviço, a atuação de agentes de atendimento treinados com inteligência artificial

generativa, a criação de células especializadas em quebra de objeções, a formação de equipes voltadas à recuperação de quebra de acordos e o direcionamento mais preciso das monitorias realizadas pelo time de qualidade. Dessa forma, a articulação entre análise descritiva, modelagem estatística e uso da informação evidencia o papel central da CI na organização, no fluxo e na aplicação estratégica das informações no processo de gestão da inadimplência.

5 Considerações Finais

Através deste estudo foi possível compreender o perfil dos inadimplentes da carteira analisada, evidenciando como a integração entre a Estatística e a CI é importante para interpretar de forma sistemática o cenário de endividamento no Brasil. A partir da análise dos dados, foi possível transformar registros operacionais em informação qualificada, gerando conhecimento aplicado para subsidiar a tomada de decisão estratégica. Essa integração mostra o como a CI é interdisciplinar, dialogando de maneira direta com áreas como Estatística, Administração, Economia, Ciência de Dados e Tecnologia da Informação, ampliando sua capacidade explicativa e de intervenção sobre problemas complexos da realidade social e organizacional. Os resultados mostram que o valor dos modelos estatísticos, como a regressão logística, não reside exclusivamente em sua capacidade preditiva, mas na qualidade semântica, organizacional e ética da informação que produzem. Em ambientes marcados pelo uso constante de dados sensíveis, como o setor de crédito e cobrança, a articulação entre Estatística e CI é essencial para assegurar decisões mais estratégicas, informadas, eficientes e socialmente responsáveis.

No contexto brasileiro, caracterizado por elevados índices de inadimplência, as organizações que atuam na concessão de crédito e na recuperação de dívidas dependem de uma complexa cadeia informacional composta por dados cadastrais, históricos de pagamento, informações comportamentais e indicadores macroeconômicos, integrando dimensões tecnológicas, organizacionais e sociais. O estudo evidencia que apenas a articulação coerente dessas dimensões permite uma gestão eficaz do risco de crédito.

Como contribuição prática, o trabalho apresenta-se como um instrumento de caráter consultivo para organizações que atuam no setor de recuperação de crédito, ao demonstrar a aplicabilidade da regressão logística como apoio à segmentação de carteiras, à priorização de contatos e à definição de estratégias de cobrança mais personalizadas. Além disso, os resultados indicam a relevância do investimento em sistemas mais sofisticados de gestão da informação, com integração de bases de dados, monitoramento contínuo de risco, automações de processos e incorporação de soluções baseadas em Inteligência Artificial, como agentes de atendimento e plataformas de autosserviço, capazes de potencializar a eficiência operacional e a qualidade do atendimento. No campo teórico, o estudo reforça o papel central da CI na transformação de dados em conhecimento, evidenciando sua contribuição para a inteligência organizacional e social em ambientes orientados por dados. Além disso, a abordagem metodológica e analítica utilizada não se restringe ao setor de cobrança, podendo ser adaptada a outros segmentos econômicos, como instituições financeiras, varejo, telecomunicações, seguros, saúde, setor público e demais organizações que operam com grandes volumes de dados e tomam decisões sob incerteza.

Por fim, para pesquisas futuras, sugere-se a ampliação do conjunto de variáveis explicativas, a incorporação de indicadores macroeconômicos mais detalhados e comparação entre a regressão logística e métodos mais avançados de aprendizado de máquina, como árvores de decisão, *random forest*s e redes neurais. Também se recomenda a realização de estudos sobre os impactos éticos, sociais e organizacionais do uso da Inteligência Artificial na gestão da inadimplência, ampliando o debate sobre o papel da CI na governança dos dados e na construção de decisões mais eficazes e responsáveis.

6 Referências Bibliográficas

BANCO CENTRAL DO BRASIL. *Recomendações de Basileia*. Disponível em: https://www.bcb.gov.br/estabilidadefinanceira/recomendacoesbasileia. Acesso em: 21 nov. 2025.

BANCO CENTRAL DO BRASIL. *Resolução CMN no 4.966, de 25 de novembro de 2021. Dispõe sobre classificação, mensuração, perdas esperadas e demais requisitos relativos a instrumentos financeiros pelas instituições autorizadas pelo Banco Central do Brasil*. Disponível em: https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolução%20CMN&numero=4966. Acesso em: 21 nov. 2025.

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.

CHOO, Chun Wei. *A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões*. Tradução: Eliana Rocha. São Paulo: Editora Senac São Paulo, 2003.

COX, D. R. *The analysis of binary data*. London: Methuen, 1970.

COSTA, Marcelo Azevedo. *Tópicos em ciências de dados: introdução dos modelos paramétricos e suas aplicações utilizando o R*. Rio de Janeiro: Bonecker, 2019.

DAVENPORT, T. H. *Ecologia da informação: porque só a tecnologia não basta para o sucesso na era da informação*. São Paulo: Futura, 1998.

DUTRA, Juliana Simões da Silva. *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.

EFROYMSON, M. A. *Multiple regression analysis. In: RALSTON, A.; WILF, H. S. (ed.). Mathematical methods for digital computers*. New York: Wiley, 1960.

FLORIDI, Luciano. *The philosophy of information*. Oxford: Oxford University Press, 2011.

GIL, Antônio Carlos. *Métodos e técnicas de pesquisa social*. 6. ed. São Paulo: Atlas, 2008.

MACHADO, Aline Rodrigues. *Collection Scoring via Regressão Logística e Modelo de Riscos Proporcionais 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: 20 nov. 2025.

SARACEVIC, Tefko. *Information science. Journal of the American Society for Information Science*, v. 50, n. 12, p. 1051–1063, 1999.

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.

SIMON, Herbert A. *Administrative behavior: a study of decision-making processes in administrative organizations*. 4. ed. New York: Free Press, 1997.

SERASA EXPERIAN. *Mapa da inadimplência e renegociação de dívidas*. Disponível em: https://www.serasa.com.br/limpa-nome-online/blog/mapa-da-inadimplencia-e-renogociacao-de-dividas-no-brasil/. Acesso em: 21 nov. 2025.

SERASA EXPERIAN. *Indicadores*. Disponível em: https://www.serasaexperian.com.br/sala-de-imprensa/indicadores/dividas-medias-de-empresas-cresceram-136-em-um-ano-revela-serasa-experian/. Acesso em: 21 nov. 2025.

SHANNON, Claude E. *A mathematical theory of communication. Bell System Technical Journal*, v. 27, p. 379–423; 623–656, 1948.

SWETS, J. A. *Measuring the accuracy of diagnostic systems. Science*, v. 240, p. 1285–1293, 1988.

[i] Descrição mais detalhada do modelo de Collection Scoring pode ser obtida em Machado (2015).
[ii] Empresa britânica de seguros fundada em 1825, mais detalhes em https://www.standardlife.co.uk.
[iii] O estatístico David R. Cox expandiu e formalizou a utilização do modelo de regressão logística, relacionando-o à família dos modelos lineares generalizados, estabelecendo o arcabouço teórico que aproximou a regressão logística ao que é utilizado atualmente.
[iv] O estatístico M. A. Efroymson propôs um algoritmo automático de seleção de variáveis em regressão múltipla no capítulo "Multiple Regression Analysis" (1960), no livro Mathematical Methods for Digital Computers, que é citado como a origem do método stepwise (incluindo a variação forward).
[v] John A. Swets é considerado o principal responsável por difundir a análise ROC na estatística moderna.
[vi] Descrição mais detalhada do conceito de razão de chances (OR – Odds ratio) pode ser obtida em Costa (2019).
[vii] Estudo divulgado pelo Serasa Experian mensalmente, que mostra o cenário de endividamento e renegociação de dívidas do Brasil. Disponível em https://www.serasa.com.br/limpa-nome-online/blog/mapa-da-inadimplencia-e-renogociacao-de-dividas-no-brasil/.