57º SBPO
SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL
Gramado • RS
DE 5 A 9 DE OUTUBRO DE 2025

Segmentação de Modelos de Elasticidade via BRKGA no Processo de Otimização de Preços de Produtos de Crédito

Guilherme Pisani[1], **Renato Cesar Sato, Antônio Augusto Chaves**
UNIFESP - Universidade Federal de São Paulo
guilherme.pisani@unifesp.br, antonio.chaves@unifesp.br, rcsato@unifesp.br

RESUMO

O processo de precificação de crédito evoluiu nos últimos anos, chegando ao estado atual, que utiliza ferramentas analíticas combinadas para realizar a otimização de preços. Neste problema, modelos de ação-e-reação são utilizados para estimar o comportamento esperado dos clientes em resposta à variação da taxa oferecida. Este trabalho visa analisar o comportamento do processo de otimização, dada a variação dos modelos de ação-e-reação utilizados para prever a probabilidade de aceitação. Nos experimentos realizados, observou-se que uma segmentação otimizada é uma etapa fundamental no processo de modelagem, identificando grupos com diferentes respostas à variação da taxa, antes do desenvolvimento de modelos estatísticos para cada grupo. Uma segmentação é proposta utilizando a metaheurística BRKGA, determinando a seleção de variáveis e a definição de agrupamentos de clientes. Os resultados mostram que o modelo de ação-e-reação desenvolvido utilizando a segmentação proposta apresenta melhor desempenho, traduzido em maior rentabilidade esperada.

PALAVRAS CHAVE. Modelo de ação-e-reação, precificação de crédito, segmentação, BRKGA.

MH – Metaheurísticas, OA – Outras Aplicações em PO

ABSTRACT

The credit pricing process has evolved over the last years, reaching the current state, which uses combined analytical tools to perform price optimization. In this problem, action-effect models must be used to estimate the expected behavior of customers in response to the variation of the offered rate. This work aims to analyze the behavior of the optimization process, given the variation of the action-effect models used to predict the probability of acceptance. In the performed experiments, it was observed that the correct segmentation is a fundamental step in the modelling process, identifying groups with different responses to the variation of the rate, before developing statistical models for each group. A novel segmentation is proposed using the BRKGA metaheuristic, determining the selection of variables and the definition of customer clusters. The results show that the action-effect model developed using the proposed segmentation presents better performance, translated as higher expected profitability.

KEYWORDS. Action-effect model, credit pricing, segmentation, BRKGA.

MH – Metaheuristics, OA – Other Applications in PO

[1]O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

DOI: 10.59254/sbpo-2025-212410
https://proceedings.science/p/212410?lang=pt-br

1. Introdução

A Pesquisa Operacional (PO) vem sendo utilizada há vários anos em diversos contextos para contribuir com a tomada de decisão dos gestores de diversas áreas. Neste cenário, cada vez mais surgem aplicações onde há a combinação da PO com outros métodos analíticos, de forma a resolver problemas complexos onde os objetivos a serem otimizados, muitas vezes, dependem da probabilidade de um evento acontecer. Para determinar a probabilidade de acontecimento destes eventos, são utilizados modelos estatísticos de aprendizado de máquinas, chamados de modelos de ação e reação, que visam prever como a variável de decisão influencia na variação desta probabilidade. Neste caso, o modelo é incorporado à função objetivo e altera as curvas originais, adicionando um efeito não linear à função.

Um exemplo desta aplicação de modelos de ação e reação é a otimização de preços, onde a alteração do preço oferecido (variável de decisão) afeta a demanda, uma vez que a disposição dos clientes de pagar por um determinado produto varia de acordo com o preço ofertado [Phillips, 2005]. Desta forma, pode-se utilizar um modelo matemático de elasticidade de preço, o qual relaciona o aumento ou diminuição na disposição dos clientes em efetuar a compra de um produto com a mudança no preço ofertado [Terblanche and De la Rey, 2014].

Para a aplicação nos processos de precificação de empréstimos, a modelagem da demanda é usualmente realizada através de um processo de aprendizado de máquinas, dada a natureza dos dados capturados no processo, que torna possível a identificação de clientes que aceitaram a oferta realizada e os que não aceitaram. O processo de aprendizado de máquinas faz referência à capacidade que os modelos têm de melhorar sua própria performance utilizando dados, sem a necessidade de ser explicitamente programado [Mitchell, 1997]. Este processo é realizado essencialmente utilizando-se modelos estatísticos que são programados para buscar melhorias através de tentativas de iterações [Gong, 2022].

Neste cenário, faz-se importante conhecer como se dá a interação entre o modelo de aprendizado de máquina e o processo de otimização, bem como quais são os fatores que podem influenciar o resultado obtido neste processo. Com o conhecimento das principais componentes que influenciam a performance dos modelos de ação e reação no processo de otimização, pode-se propor melhorias no desenvolvimento dos modelos, buscando obter resultados superiores na otimização de preços.

Sendo assim, este trabalho propõe uma análise do processo de precificação de produtos de crédito, avaliando o resultado de receita esperada ao final do processo de otimização, para encontrar os componentes que podem ser alterados para a obtenção de um melhor resultado. Para isto, o processo completo de desenvolvimento da precificação será implantado, para que se possa medir qual o impacto observado pelas mudanças no resultado ao final da otimização. O principal enfoque será nas etapas de segmentação e desenvolvimento dos modelos de ação e reação, para verificar como eles alteram o resultado e identificar pontos de melhoria. É proposto um método de segmentação da elasticidade de preços, baseado na metaheurística BRKGA [Gonçalves and Resende, 2011], de forma a encontrar grupos que apresentem diferentes reações à mudança do preço, para posterior desenvolvimento dos modelos de aprendizado de máquina para cada um dos grupos.

Este trabalho está organizado da seguinte forma. A seção 2 apresenta o problema abordado. A seção 3 descreve a metodologia utilizada nesta pesquisa, apresentando a criação das bases de dados, os modelos de risco e aceitação e a metaheurística BRKGA para a segmentação dos clientes. A seção 4 apresenta os resultados computacionais e na seção 5 faz-se as considerações finais.

DOI: 10.59254/sbpo-2025-212410
https://proceedings.science/p/212410?lang=pt-br

2. Definição do Problema

O mercado de crédito exerce uma grande influência na economia, trazendo impactos significativos no nível de atividade e no poder de compra da população. A correta gestão dos riscos envolvidos neste tipo de operação é fundamental para garantir a estabilidade econômica. Entre as variáveis utilizadas pelas instituições para a gestão da carteira de crédito, encontra-se a precificação, determinada pela taxa de juros cobrada de cada cliente, que contribui para o atingimento dos objetivos definidos pela instituição, como a rentabilidade desejada, participação de mercado e controle de risco.

A aplicação da otimização de preços permite o ganho de flexibilidade e agilidade, aliado à busca de soluções otimizadas, que resultem em melhores decisões, incrementando o resultado das instituições. Nesta classe de otimização, faz-se necessária a utilização de modelos de aprendizado de máquina como parte integrante do processo, trazendo uma componente extra que necessita ser desenvolvida e monitorada.

O *Risk-based Pricing* (precificação baseada no risco) surgiu como uma resposta à ineficiência do sistema utilizado na década de 90, no qual todos os clientes recebiam a mesma taxa. Esse novo modelo ajusta as taxas de acordo com o risco de cada cliente, permitindo que aqueles com maior probabilidade de inadimplência paguem taxas mais altas para compensar o risco associado a possíveis perdas, e foi impulsionado pela melhora dos modelos de classificação de risco e pela redução dos custos de armazenamento de dados [Phillips, 2013; Bostic, 2002].

Durante um longo período, esta foi a maneira que a maioria das instituições construía seus modelos de precificação e que ainda continua a ser utilizada em muitas instituições [Edelberg, 2006; Chun and Lejeune, 2020]. Apesar de proporcionar um avanço frente ao modelo de precificação única, precificação baseada em risco apresenta limitações, sendo a mais notável a falta de incorporação da propensão dos clientes a aceitar a oferta, dado o preço [Chun and Lejeune, 2020]. A falta desta informação no modelo traz consequências ao resultado da instituição, uma vez que ignora as diferentes características dos clientes em relação à variação da taxa, não permitindo a visualização completa do resultado da escolha das diferentes taxas para cada segmento de clientes.

A precificação baseada no risco é a versão da indústria financeira para o preço baseado em custos, no qual o preço de um produto ou serviço leva em conta o custo mais uma margem de lucro fixa, e se faz importante uma abordagem de precificação baseada em lucro, abordagem esta que engloba tanto a visão do risco quanto a variação da demanda, de forma a maximizar o lucro, conceito este semelhante ao utilizado com sucesso em outras indústrias, como hotéis, companhias aéreas, telecomunicações e varejo [Caufield, 2012; Phillips, 2005].

Uma vantagem da indústria financeira para a utilização desta informação é o fato de esta diferenciação na elasticidade observada entre os públicos poder ser mensurada, à medida que os clientes aprovados para um financiamento podem optar entre contratar ou não contratar tal financiamento, gerando um indicador conhecido por taxa de efetivação ou taxa de aceitação, calculado como: quantidade de financiamentos contratados / quantidade de financiamentos aprovados. Através deste indicador de taxa de efetivação, temos a referência do percentual de clientes que de fato contratam o financiamento naquele segmento e é possível armazenar os dados de sucessos e insucesso de vendas em base de dados estruturada, permitindo a modelagem do evento [Renison and Hardt, 2008; Dey, 2010]. Dada esta característica de disponibilidade de informações, a modelagem da probabilidade de aceitação de empréstimos pode ser realizada com técnicas de aprendizado de máquina, sendo tratada como um problema de classificação, que será objeto de estudo deste trabalho.

DOI: 10.59254/sbpo-2025-212410
https://proceedings.science/p/212410?lang=pt-br

Muitas instituições seguem um macroprocesso padrão de concessão de crédito, com uma sequência de análises e decisões que resultam na aprovação e precificação do crédito individualizado de cada solicitação. Este processo se inicia com o preenchimento e envio da proposta pelo cliente, e é seguido de uma busca de dados internos e externos que são utilizados para calcular a probabilidade de inadimplência da solicitação, composta por uma combinação entre o risco do cliente e da operação analisada, que é utilizada para balizar a decisão de aprovação ou negação da proposta. Para os clientes aprovados no processo de avaliação creditícia, são definidas das taxas de juros (precificação), processo esse que é escopo deste estudo. Ao final, o cliente tem a decisão final se aceita ou não o crédito, dadas as condições oferecidas. A Figura 1 ilustra este macroprocesso de concessão.

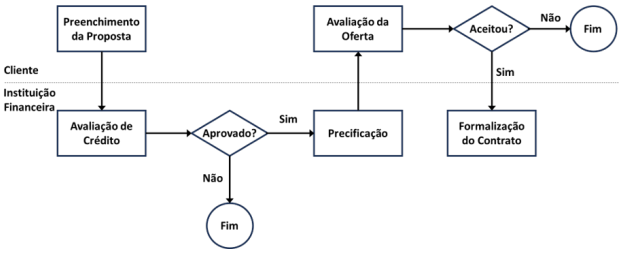


Figura 1: Macrofluxo do processo de concessão de crédito.

Nos processos mais avançados de definição da estratégia de precificação, são utilizados os modelos de aceitação e a otimização de preços. As etapas para a definição da estratégia são apresentadas na Figura 2 e consistem em, a partir de uma base de dados histórica, criar uma segmentação de clientes, desenvolver os modelos de aceitação para cada um dos segmentos, aplicar o processo de otimização de preços e analisar o resultado de receita esperada.

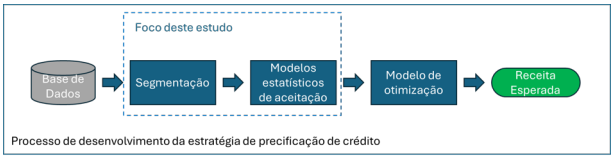


Figura 2: Processo de desenvolvimento da estratégia de precificação.

O alvo deste estudo será a alteração e simulação de diferentes formas de segmentação e modelagem de aceitação, avaliando não somente os desempenhos estatísticos de separação dos modelos, mas principalmente o resultado da receita esperada após a otimização de preços, que é o objetivo final de todo este processo.

3. Metodologia

O presente trabalho foi desenvolvido através de uma pesquisa exploratória composta, de forma resumida, pelas seguintes etapas: 1) criação de uma base de dados sintética simulando uma base de concessão de crédito veicular parcelado; 2) desenvolvimento do modelo de risco para estimação da probabilidade de inadimplência; 3) desenvolvimento de uma proposta de segmentação de públicos, baseada na elasticidade da demanda, usando um algoritmo BRKGA; 4) desenvolvimento de diversos modelos estatísticos de aceitação, com diferentes conceitos de segmentação e combinações de variáveis para prever a probabilidade de aceitação do cliente de acordo com a taxa ofertada; 5) implementação de um modelo

DOI: 10.59254/sbpo-2025-212410
https://proceedings.science/p/212410?lang=pt-br

de otimização de taxa de juros, com o objetivo de maximizar a rentabilidade, utilizando-se o modelo de aceitação, de forma a estimar a receita esperada; 6) execução do processo de otimização com os diversos modelos probabilísticos, permitindo a comparação entre o desempenho individual dos diferentes modelos e os respectivos resultados obtidos no processo de otimização, através da receita esperada.

3.1. Criação da Base de Dados

Dadas as novas regulamentações sobre a proteção de dados cadastrais, nos últimos anos as instituições financeiras se tornaram receosas em compartilhar informações de clientes, por risco de terem sua imagem ligada ao possível vazamento de dados e às diversas implicações que isso pode ter, como por exemplo, multas e perda de clientes. Desta forma, neste trabalho foi gerada uma base de dados sintética, utilizando parâmetros de variáveis conhecidas e seus respectivos poderes de explicação para os eventos de perda e aceitação. Como o presente estudo tem como objetivo avaliar o desempenho relativo entre os modelos, a base sintética atende a este requisito, uma vez que a comparação sempre se dará com um cenário base também desenvolvido sobre a mesma amostra. Vale ressaltar que, uma vez criada, a base é estática, e as únicas alterações feitas são nos modelos que serão testados, sempre utilizando o mesmo conjunto de dados. Por isso, a abordagem permite comparar o efeito destas alterações. Esta base está disponível em `https://www.kaggle.com/datasets/guilhermepisani16/autoloan-risk-and-pricing`.

3.2. Desenvolvimento do Modelo de Risco

O desenvolvimento do modelo de risco utiliza a técnica de regressão logística para a determinação da probabilidade de inadimplência de cada cliente e a criação de cinco níveis de risco que posteriormente serão utilizados no processo de otimização.

Para todos os modelos de regressão logística gerados, tanto de risco quanto de aceitação, a base é separada em amostra de treino e amostra de validação, com uma proporção de 70% e 30%, respectivamente, e posteriormente rodada uma regressão logística no Python, utilizando-se do pacote statsmodels. O modelo de risco gerado apresentou resultados de separação satisfatórios e muito semelhantes entre as amostras de treino e validação (Treino: AUC de 85,3% e KS de 54,2%; Validação: AUC de 85,3% e KS de 54,3%). Com a informação da probabilidade de inadimplência foram criados 5 grupos de nível de risco, contendo aproximadamente 20% da população em cada grupo.

3.3. Desenvolvimento dos Modelos de Aceitação

Para a modelagem de aceitação, foi realizado um processo de seleção de variáveis através do indicador *Information Value* (IV), a partir do qual foram selecionadas as variáveis comprometimento de renda, nível de risco, escolaridade e taxa de juros. Basicamente existem três componentes que serão analisadas e manipuladas durante o desenvolvimento dos modelos, sendo elas: 1) Desempenho do modelo estatístico, medida pelo AUC, que será manipulada através da utilização de mais ou menos variáveis; 2) Segmentação utilizada para definir diferentes públicos a serem modelados de forma independente, por apresentarem respostas diferentes à variação da taxa de juros, medida pela diferença dos coeficientes lineares da função aceitação pela taxa de juros de cada segmento (este conceito será exemplificado na seção 3.5); 3) Resultado da receita esperada obtido ao final do processo de otimização em relação à performance estatística de cada modelo, medida pelo AUC.

São desenvolvidos oito modelos de aceitação com diferentes abordagens, usando as variáveis pré-selecionadas. O Modelo 1 utiliza um único ajuste para toda a base com as variáveis comprometimento de renda, nível de risco e taxa como preditoras. O Modelo 2 segmenta a base por níveis de risco, baseado no conceito de seleção adversa, ajustando

DOI: 10.59254/sbpo-2025-212410
https://proceedings.science/p/212410?lang=pt-br

uma regressão logística para cada grupo, usando comprometimento de renda e taxa como preditoras. O Modelo 3 cria oito grupos combinando comprometimento de renda e risco, pois estas variáveis são as que melhor separam grupos com elasticidades distintas, modelando cada grupo apenas com a variável taxa. No Modelo 4, são mantidos os grupos do Modelo 3, mas adiciona-se comprometimento de renda e nível de risco como preditoras para melhorar o ajuste. O Modelo 5 ajusta um único modelo usando apenas nível de risco e taxa, enquanto o Modelo 6 segmenta a base por nível de risco, utilizando somente a taxa como preditora. O Modelo 7 aplica uma regressão única para toda a base considerando apenas a taxa. Por fim, o Modelo 8 utiliza a segmentação otimizada obtida por meio da metaheurística BRKGA (seção 3.5), mantendo as mesmas variáveis preditoras dos modelos anteriores. Todos os modelos são gerados em uma amostra de treino e verificados em uma amostra de validação, apresentando performances similares, indicando não haver *overfitting*.

3.4. Modelo de Otimização

Dado o objetivo de avaliar o resultado através da receita obtida no processo de otimização, comparando os diferentes modelos de aceitação, foi implementado um modelo de otimização de referência da literatura, proposto por [Terblanche and De la Rey, 2014], no qual utilizamos as informações dos modelos de aceitação - *T(r)* - com a aproximação linear da receita esperada - *I(r,a,n,p)*, representando a fórmula da receita esperada para cada cliente como a multiplicação das duas componentes, conforme Equação 1.

$$\text{Receita do Portfolio} = \sum_{c \in C} T(r_c) * I(r_c, a_c, n_c, p_c) \tag{1}$$

sendo: C o conjunto de clientes, r_c o *spread* mensal, a_c o valor concedido, n_c o número de parcelas (prazo em meses), p_c a probabilidade de inadimplência, δ o percentual de perda dada a inadimplência (LGD = 80%) e $I = n_c.a_c.r_e - a_c.p_c.\delta$

Como a curva de aceitação $T(r_c)$ é uma função logística, observamos que tanto esta função, como a multiplicação da aproximação linear da receita $I(r_c, a_c, n_c, p_c)$ com a probabilidade de aceitação $T(r_c)$, são funções não lineares. Com o objetivo de simplificar a resolução deste problema, [Terblanche and De la Rey, 2014] propõem a utilização da linearização *piece-wise*, criando segmentos de retas que serão utilizados no processo de otimização, fazendo com que o problema possa ser resolvido pelos *solvers* lineares.

Utilizando o modelo proposto no trabalho de [Terblanche and De la Rey, 2014] e aplicando a linearização *piece-wise*, assumimos $0 \leq r_c \leq 1$, sendo dividido em intervalos indexados por $K = 1, 2, ..., 1/w$, sendo $0 \leq w \leq 1$ o tamanho do intervalo. Para cada intervalo $k \in K$, a resposta da função de probabilidade de aceitação é aproximada pela função linear da *piece-wise*, e a inclinação (*slope*) desta função é definida como m^T_{ck}.

Da mesma forma, é realizada aproximação através da linearização *piece-wise* da função objetivo $T(r_c) * I(r_c, a_c, n_c, p_c)$ (Rentabilidade do Portfólio), sendo a inclinação da reta em cada segmento denominada como m^I_{ck}.

São introduzidas também variáveis incrementais $0 \leq r_{ck} \leq w$, que é a taxa aplicada em cada segmento, sendo ela um valor entre zero e o tamanho máximo do segmento. Importante observar que a taxa que será atribuída a cada cliente será a soma de todos os r_k deste cliente, pois esta é uma variável incremental, considerada apenas para cada segmento de reta da *piece-wise*. Foi introduzida também a variável t_c, que é a probabilidade de aceitação de cada cliente c. Utilizando estes conceitos, foi definido um modelo linear de otimização da taxa de juros:

$$maximizar \sum_{c \in C} \sum_{k \in K} m^I_{ck}.r_{ck} \tag{2}$$

DOI: 10.59254/sbpo-2025-212410
https://proceedings.science/p/212410?lang=pt-br

Sujeito à:

$$t_c - \sum_{k \in K} m_{ck}^T = T_c(0) \quad \forall c \in C \tag{3}$$

$$t_c \geq 0 \quad \forall c \in C \tag{4}$$

$$0 \leq r_{ck} \leq w \quad c \in C, k \in K \tag{5}$$

A Restrição 3 é inserida para que seja possível atribuir valores à variável t_c, que representa a probabilidade de aceitação para a taxa ótima, que é comumente utilizada para criar uma restrição de volume mínimo de concessão, de forma a maximizar a rentabilidade da carteira sem perder volume e, consequentemente, participação de mercado. Nesta restrição encontramos a variável $T_c(0)$, que representa a taxa de aceitação de cada cliente para a taxa com valor zero. A Restrição 4 garante que a aceitação seja maior ou igual a zero, enquanto a Restrição 5 delimita o valor das variáveis incrementais r_{ck} dentro dos intervalos da *piece-wise*.

3.5. Modelo de Segmentação Utilizando BRKGA

Como será apresentado na seção de resultados deste trabalho, a segmentação de clientes, de forma a identificar grupos com diferentes sensibilidades à mudança de preços, mostrou-se um fator crucial durante o desenvolvimento dos modelos de aceitação visando a obtenção de bons resultados no processo de otimização.

Para determinar qual é a sensibilidade ao preço de cada segmento, foi utilizado como indicador a inclinação da reta resultante da regressão linear dos pontos observados de aceitação para cada uma das taxas existentes. A Tabela 1 exemplifica como são apresentados os pontos de aceitação para dois segmentos hipotéticos de clientes, de acordo com as taxas oferecidas. Observa-se que conforme a taxa aumenta, um menor percentual de clientes aceita a oferta. Com esta informação, uma regressão linear pode ser calculada e o coeficiente linear Alfa (Equação 9) é utilizado para determinar a inclinação da reta, que representa a sensibilidade deste grupo à alteração da taxa de juros oferecida.

Tabela 1: Percentual de aceitação por taxa para cálculo da inclinação.

	Taxa					
	1.1	1.3	1.5	1.7	1.9	Inclinação
Segmento 1	72%	68%	58%	45%	40%	-43.5%
Segmento 2	57%	54%	48%	41%	37%	-26.5%

Para maximizar a diferença dos Alfas de cada segmento no processo de segmentação, foi utilizada a metaheurística *Biased Random-Key Genetic Algorithm* (BRKGA) [Gonçalves and Resende, 2011] para resolver o problema de agrupamento. O BRKGA é um algoritmo derivado dos Algoritmos Genéticos que tem sido usado com sucesso em vários problemas clássicos de otimização combinatória classificados como NP-difícil, destacando-se sobre outras variantes de algoritmo genético pela rápida convergência para soluções de alta qualidade [Andrade et al., 2021].

O BRKGA caracteriza-se por ser um algoritmo genético que utiliza vetores de números entre zero e um (chaves) para representar soluções. O processo do BRKGA consiste basicamente nas seguintes etapas: 1) Inicialização: Geração aleatória de uma população inicial, representada pelos vetores de chaves; 2) Decodificação: Conversão do vetor de chaves em soluções viáveis por meio do *decoder*, chegando ao custo de cada solução; 3) Seleção:

DOI: 10.59254/sbpo-2025-212410
https://proceedings.science/p/212410?lang=pt-br

Classificação da população através do custo em indivíduos elite e não-elite; 4) Reprodução: Evolução de geração, transferindo os indivíduos elite e gerando novos indivíduos através do cruzando de indivíduos elite com não-elite, com viés favorecendo genes da elite; 5) Mutação: Inclusão de novos indivíduos aleatórios para manter diversidade e tentar evitar soluções ótimas locais; 6) Evolução: O processo se repete até um critério de parada ser atingido.

A proposta do BRKGA foi desenvolvida com o objetivo de resolver dois problemas de forma concomitante, sendo estes problemas a seleção das características (variáveis) que serão utilizadas para a segmentação, e o segundo, a definição de em qual grupo cada cliente estará. Para isto, foi desenvolvido um *decoder*, que faz a conversão de uma solução de chaves aleatórias (valores reais no intervalo $[0, 1)$) em uma solução do problema específico que está sendo tratado, de forma a gerar uma solução factível do problema e seu respectivo custo.

As soluções são codificadas como vetores n-dimensionais, sendo que n é o número de características (variáveis) disponíveis na base de dados. O *decoder* implementado neste trabalho realiza a seleção das características que serão utilizadas na segmentação e aplica uma heurística para realizar os agrupamentos de clientes e definir os *clusters*.

A primeira etapa do *decoder* consiste em transformar os valores da solução de chaves aleatórias em valores binários, alocados em uma variável X. A conversão é realizada da seguinte maneira: se o valor da chave aleatória for menor que 0,5, a característica v correspondente não é selecionada ($X_v = 0$); caso contrário, a característica é selecionada ($X_v = 1$). Dessa forma, o conjunto de características escolhidas para compor a segmentação é definido com base nessa classificação binária.

Todas as características utilizadas foram categorizadas em, no máximo, nove categorias e transformadas em um campo numérico, denominado CL. Na sequência, multiplicamos este campo pela variável X, obtida na etapa 1 do *decoder*. Como resultado desta multiplicação, temos uma nova variável, chamada de XCL, descrita na Equação 6.

$$XCL_{vc} = X_v * CL_{vc} \quad \forall c \in C, v \in V \tag{6}$$

Posteriormente, as informações contidas no campo XCL são consolidadas em um único campo, chamado de identificador, CH, através do cálculo realizado na Equação 7 e que tem a finalidade de garantir que clientes com as mesmas características possuam o mesmo identificador.

$$CH_c = \sum\nolimits_{v \in V} XCL_{vc} * 10^{v-1} \quad \forall c \in C \tag{7}$$

Neste problema existe uma restrição de alocação de clientes nos grupos que deve ser respeitada, garantindo que clientes com as mesmas características fiquem necessariamente no mesmo grupo, motivo pelo qual foi utilizado o conceito de uma restrição do tipo *Must-Link* ($\mathcal{ML}$). A $\mathcal{ML}$ é utilizada em técnicas de clusterização, onde os pontos que devem estar necessariamente em um mesmo grupo são colocados aos pares na matriz $\mathcal{ML}$.

Dada a natureza do problema, há um desafio para o cumprimento desta restrição, devido ao fato de a matriz $\mathcal{ML}$ somente ser conhecida em tempo de execução, pois ela depende diretamente das características que serão utilizadas para realizar a segmentação. Por este motivo, foi utilizada a abordagem de criação do identificador CH, que após ser calculado, passa a ser utilizado para o agrupamento dos registros, garantindo o cumprimento da restrição do tipo $\mathcal{ML}$.

Após este agrupamento, é contada dentro de cada CH e de cada Taxa, a quantidade total de registros ($\sum QG$) e a quantidade de registros efetivados ($\sum EG$) e através da divisão

DOI: 10.59254/sbpo-2025-212410
https://proceedings.science/p/212410?lang=pt-br

do segundo pelo primeiro, obtém-se a taxa de efetivação (EFT), separada pelas dimensões de grupo e taxa oferecida, Equação 8. Com esta informação, é possível calcular o valor dos Alfas de cada subgrupo de identificador, que demonstra a variação da resposta, dada a variação da taxa, Equação 9. Na sequência, os subgrupos são ordenados de forma crescente pelos Alfas e os grupos provisórios são atribuídos em uma divisão uniforme da quantidade de identificadores pela quantidade de grupos.

$$EFT_{gt} = \frac{\sum_{c \in C} EG_{gtc}}{\sum_{c \in C} QG_{gtc}} \quad \forall g \in G, t \in T \tag{8}$$

$$Alfa_g = \frac{q_t * \sum_{t \in T} EFT_{gt} * T_t - \sum_{t \in T} EFT_{gt} * \sum_{t \in T} T_t}{q_t * \sum_{t \in T} T_t^2 - (\sum_{t \in t} T_t)^2} \quad \forall g \in G \tag{9}$$

Após esta definição, a heurística analisa os identificadores limítrofes dos grupos provisórios, verificando se os clientes que estão na fronteira são mais parecidos com o grupo atual ou com o adjacente, comparando a diferença dos *Alfas*, alocando o cliente analisado no grupo mais parecido. Esta lógica é aplicada de forma iterativa, até que não haja mais mudanças para serem realizadas. Ao final, cada identificador será atribuído a um *cluster*, fazendo com que a restrição de *mustlink* seja cumprida. Para evitar soluções que contenham identificadores com um pequeno número de clientes e, portanto, sem relevância estatística, foi incluída uma condição de quantidade mínima de registros, denotada por m, gerando uma penalização na solução final. Esta condição também é útil para limitar a quantidade de variáveis selecionadas, pois quanto mais variáveis são utilizadas para a segmentação, maior a chance de haver grupos muito pequenos.

O custo da solução então é calculado através da somatória da diferença dos *Alfas* dos grupos, onde o objetivo do problema é a maximização destas diferenças, de forma a encontrar a segmentação que apresente grupos mais distantes entre si com relação à resposta da aceitação de acordo com a taxa ofertada. A Equação 10 apresenta o objetivo do problema.

$$Max \sum_{g \in G} (Alfa_g - Alfa_{g+1}) \tag{10}$$

As Figuras 3 e 4 ilustram a lógica desenvolvida neste *decoder*, construída com base nas equações e na heurística previamente descritas.

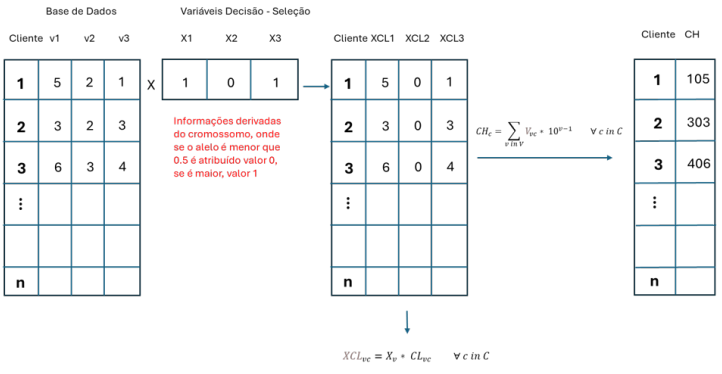


Figura 3: Exemplo de aplicação do *decoder* - etapa 1.

DOI: 10.59254/sbpo-2025-212410
https://proceedings.science/p/212410?lang=pt-br

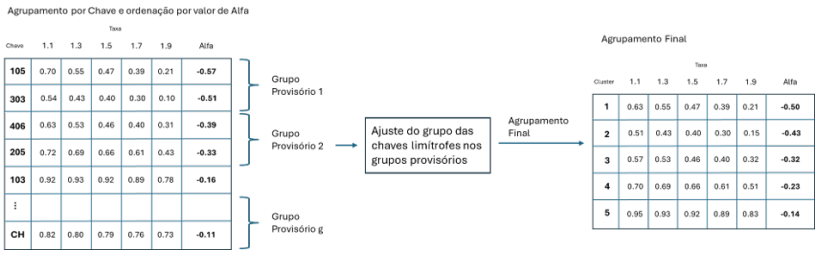


Figura 4: Exemplo de aplicação do *decoder* - etapa 2.

4. Resultados Computacionais

Com o objetivo de avaliar o resultado que cada modelo desenvolvido é capaz de produzir ao final do processo de otimização, os oito modelos desenvolvidos foram testados e organizados em um gráfico, apresentado na Figura 5, onde pode-se observar a receita esperada de cada modelo em relação à sua performance estatística, representada pelo AUC.

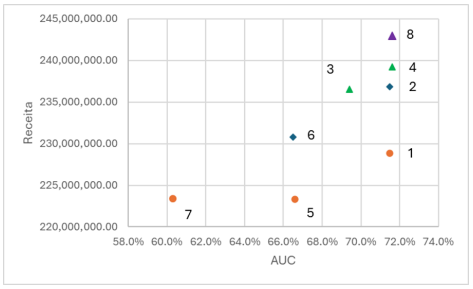


Figura 5: Comparativo da performance dos modelos com seus respectivos resultados de receita na otimização.

Os resultados observados apontam que não parece haver uma relação direta entre o poder de separação do modelo estatístico com o resultado que estes modelos conseguem entregar no processo de otimização. Nota-se que existem modelos que apresentam performances até piores, mas que alcançam melhores resultados na otimização. Sempre quando observada esta diferença de performance, percebe-se que a melhora está mais ligada à segmentação dos públicos.

Como exemplos destas observações, podemos citar a comparação entre os modelos 1 e 6, sendo que o modelo 1 utiliza mais variáveis, apresenta performance superior, porém entrega um resultado na otimização mais baixo que o modelo 6, que por sua vez é um modelo com menos variáveis, performance pior, porém segmentado, permitindo melhor separação da elasticidade e, consequentemente, melhor resultado na otimização.

Da mesma maneira, comparando os modelos 2 e 4, e considerando que ambos são segmentados e apresentam poder de separação muito similar, vemos uma vantagem no resultado de receita para o modelo 4, que utilizou a inclinação como critério de segmentação, apresentando amplitude máxima de 34,9%, enquanto o modelo 2 apresentou amplitude de 21,1%.

Para a geração do Modelo 8, foi realizada a segmentação pelo BRKGA, utilizando o *decoder* desenvolvido e o *framework* da metaheurística BRKGA desenvolvido por [Andrade et al., 2021], que pode ser acessado em `https://github.com/ceandrade/brkga\_mp\_ipr\_python`, configurada com os seguintes parâmetros: tamanho da população = 50, percentual

DOI: 10.59254/sbpo-2025-212410
https://proceedings.science/p/212410?lang=pt-br

de elite = 0,30, percentual de mutação = 0,15, número de pais elite = 2, total de pais = 3, função de fiés = LOGINVERSE, número de populações independentes = 3. Os testes computacionais foram realizados em um computador com processador Intel(R) Core(TM) i7-1365U de 1,8 GHz, 16 GB de memória RAM, rodando o sistema operacional Windows 11 de 64 bits. Após diversas execuções do processo com diferentes parâmetros, observou-se a convergência para boas soluções em aproximadamente 41 minutos, utilizando a configuração com uma penalização de três pontos percentuais para cada subgrupo de identificadores que contivesse menos de 100 clientes, em uma base que possui aproximadamente 110.000 registros.

Neste processo, foram selecionadas três variáveis para a geração dos agrupamentos, sendo elas: o Comprometimento de Renda, o Nível de Risco e a Escolaridade; foram encontrados oito grupos de clientes com inclinações distintas entre eles, resultando em uma amplitude máxima de 31,5% entre o grupo 1 e o grupo 8, apresentada na Tabela 2.

Tabela 2: Segmentação de clientes usando BRKGA.

	Taxas					
Grupo Segmento	1.1	1.3	1.5	1.7	1.9	Inclinação
1	54.9%	47.1%	40.9%	31.1%	14.8%	-48.1%
2	56.4%	52.5%	45.2%	37.2%	22.4%	-41.8%
3	66.2%	60.1%	56.1%	48.2%	35.3%	-36.9%
4	63.2%	61.4%	55.1%	47.2%	36.9%	-33.5%
5	71.8%	69.1%	65.5%	58.4%	48.6%	-28.5%
6	77.2%	74.1%	70.1%	65.9%	55.7%	-25.7%
7	80.6%	79.4%	75.0%	71.9%	61.8%	-22.6%
8	82.7%	80.4%	79.2%	75.4%	68.6%	-16.6%

Foi gerada uma regressão para cada grupo, resultando no modelo que apresentou o maior resultado de receita esperada, sugerindo a validade da utilização desta técnica de segmentação para a obtenção de melhores resultados da modelagem para otimização de preços. Este efeito ocorre, apesar de a amplitude máxima de inclinação encontrada na segmentação ser levemente mais baixa que no Modelo 4. Porém, quando avaliada a distribuição de clientes nos grupos, percebe-se que a segmentação do BRKGA conseguiu encontrar grupos mais uniformes, principalmente quando observados os grupos 7 e 8, que apresentam baixa elasticidade. As distribuições de clientes dos grupos gerados pela segmentação manual e pelo BRKGA são apresentadas na Figura 6, onde fica claro que os grupos 7 e 8 apresentam pouco representatividade na segmentação manual, o que não ocorre na segmentação obtida pelo BRKGA.

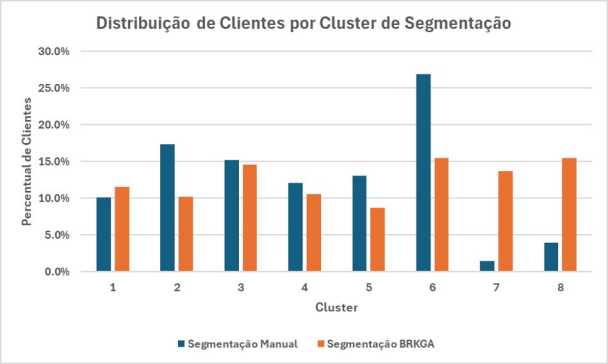


Figura 6: Comparativo da performance dos modelos com seus respectivos resultados de receita na otimização.

DOI: 10.59254/sbpo-2025-212410
https://proceedings.science/p/212410?lang=pt-br

5. Conclusão

A não relação direta entre a performance do modelo e esta aparente dependência da segmentação de públicos pode ser derivada do fato de a performance do modelo AUC ser uma medição estática, apenas conseguindo capturar o poder de segmentação do modelo para os parâmetros já definidos e fixos que se encontram na amostra analisada. Já no processo de otimização existe um campo de busca da variável resposta, fazendo com que seja inserida uma componente dinâmica relacionada à taxa de juros. Para que esta busca seja otimizada, o modelo precisa identificar esta dinâmica distinta entre os diferentes grupos da amostra, representada pela inclinação da resposta à taxa, de forma a refletir no resultado a melhor relação entre variação da taxa e resposta do modelo.

Assim, os resultados obtidos reforçam a importância de identificar públicos com comportamentos distintos em relação à variação da aceitação frente às mudanças na taxa ofertada, evidenciando que essa segmentação é crucial para o sucesso do processo de otimização. Nesse contexto, o algoritmo proposto, baseado no BRKGA, demonstrou ser uma alternativa eficaz para encontrar segmentações de alta qualidade.

Referências

Andrade, C. E., Toso, R. F., Gonçalves, J. F., and Resende, M. G. (2021). The multi-parent biased random-key genetic algorithm with implicit path-relinking and its real-world applications. *European Journal of Operational Research*, 289(1):17–30.

Bostic, R. W. (2002). Trends in equal access to credit products. In *The Impact of Public Policy on Consumer Credit*, p. 171–208. Springer.

Caufield, S. (2012). 138 Consumer Credit Pricing. In *The Oxford Handbook of Pricing Management*. Oxford University Press.

Chun, S. Y. and Lejeune, M. A. (2020). Risk-based loan pricing: Portfolio optimization approach with marginal risk contribution. *Management Science*, 66(8):3735–3753.

Dey, S. (2010). Credit limit management using action-effect models. In *2010 International Conference on Financial Theory and Engineering*, p. 112–115. IEEE.

Edelberg, W. (2006). Risk-based pricing of interest rates for consumer loans. *Journal of monetary Economics*, 53(8):2283–2298.

Gonçalves, J. F. and Resende, M. G. (2011). Biased random-key genetic algorithms for combinatorial optimization. *Journal of Heuristics*, 17(5):487–525.

Gong, M. (2022). *Studies on Classification Problems and Application in Consumer Lending Optimization System*. PhD thesis, University of Delaware, USA.

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.

Phillips, R. (2013). Optimizing prices for consumer credit. *Journal of Revenue and Pricing Management*, 12:360–377.

Phillips, R. L. (2005). *Pricing and revenue optimization*. Stanford university press.

Renison, K. and Hardt, L. (2008). Price and revenue optimization for banking. In *Proceedings of the SAS Global Forum 2008*, p. 1–7. Citeseer.

Terblanche, S. and De la Rey, T. (2014). Credit price optimisation within retail banking. *ORiON*, 30(2):85–102.

DOI: 10.59254/sbpo-2025-212410
https://proceedings.science/p/212410?lang=pt-br