

MODELAGEM E PREVISÃO DE *CHURN* EM MODELOS DE NEGÓCIO DE RECORRÊNCIA UTILIZANDO REDES NEURAIS ARTIFICIAIS

Tayrone Duque Esteves¹, Charles Dayan Farias de Jesus ², José Henrique de Freitas Gomes¹.

¹Universidade Federal de Itajubá/Programa de Pós-graduação em Engenharia de Produção, Avenida BPS, 1303, Pinheirinho – 37500-903 – Itajubá-MG, Brasil, tayrone.esteves@gmail.com, ze_henriquefg@unifei.edu.br.

² Universidade Federal de São Carlos/Departamento de Engenharia Química, Rodovia Washington Luiz, km 234 – SP 310, Monjolinho – 13565-905 – São Carlos-SP, Brasil, cdfcdf2@yahoo.com.

Resumo

O crescimento do mercado de compras recorrente tem chamado a atenção de cada vez mais empresas. Considerando o impacto que cancelamentos de assinaturas (*churn*) tem neste modelo de mercado, a retenção de clientes mostra-se essencial para o crescimento financeiro deste ramo. Nesse contexto, modelos de previsão de *churn* apresentam-se como um diferencial estratégico, pois conseguem simular o processo de tomada de decisão dos assinantes e orientar a construção de campanhas de retenção mais eficientes. Modelos regressivos, técnicas de *clustering* e redes neurais artificiais (RNA) são métodos comumente usados, individualmente ou combinados a outras ferramentas, para melhorar a acurácia das previsões de cancelamento. Neste trabalho, foram construídos dois modelos para previsão de *churn* em um clube de assinaturas que atua no mercado nacional. No primeiro aplicou-se o algoritmo média-K para dividir a amostra em 2 sub-grupos aos quais foi aplicada individualmente a regressão logística. No segundo modelo, foram utilizados os mesmos dados para treinar uma rede neural *multilayer perceptron* (MLP). Apesar do modelo composto apresentar uma precisão superior à rede MLP (55,9% e 51,7%, respectivamente), a rede neural apresentou melhor *recall* (38,3%), medida - F1 (44,0%) e retorno financeiro esperado (R\$ 57.528) mostrando uma capacidade superior de identificação de cancelamentos verdadeiros desta metodologia.

Palavras-chave: Previsão de *churn*. Redes neurais. Clubes de assinatura. Impacto econômico.

Área do Conhecimento: Engenharias – Engenharia de Produção.

Introdução

O mercado de compras recorrentes tem crescido de forma significativa nos últimos anos. Somente o mercado de clubes de assinatura teve um faturamento estimado de US\$ 2,6 bilhões no mercado americano em 2016 chegando a US\$ 28 bilhões globalmente em 2022 (Chen et al., 2018; IMARC, 2023). Este cenário altamente competitivo traz consigo a necessidade do entendimento do perfil e dos fatores que impactam a decisão de compra dos consumidores principalmente quando se considera que o custo de aquisição de novos clientes pode ser até cinco vezes superior ao custo de retenção dos atuais (Bahnsen et al., 2015).

Zhang, Moro e Ramos (2022) definem o fim do relacionamento de um cliente com uma empresa como *churn* e destacam o impacto negativo que a perda de assinantes tem no faturamento de empresas no mercado de compra recorrente. Neste cenário, a construção de modelos capazes de prever possíveis cancelamentos com antecedência torna-se um diferencial estratégico para empresas e permite a construção de campanhas de retenção mais eficientes e com melhores taxas de retenção (De Caigny et al., 2018).

Existem diversas metodologias propostas para a previsão de *churn* na literatura, desde a aplicação de métodos individualmente até a construção de processos compostos de diversas etapas e ferramentas concatenadas. As Redes Neurais Artificiais (RNA) têm se destacado nos últimos anos, principalmente pela sua capacidade de simular relações não lineares entre as variáveis preditoras e processar grandes quantidades de dados de maneira eficiente, emergindo como ferramentas de inteligência artificial para o desenvolvimento de modelos da “Indústria 4.0” (Bianchesi et al., 2022). Adicionalmente, a combinação métodos regressivos, como a regressão logística, e *clustering*, como

A Ciência do NANO e seu impacto transformador no MACRO

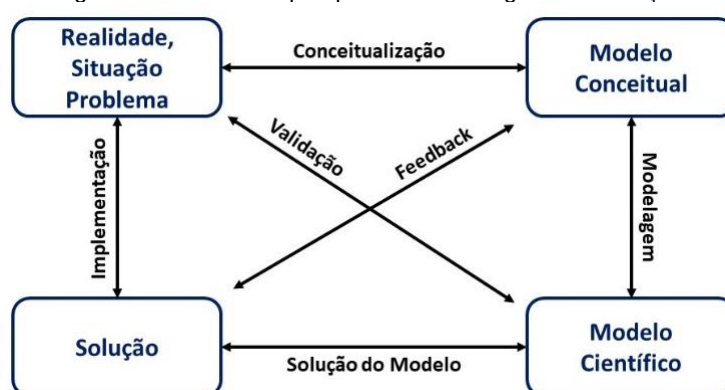
média-K, tem mostrado resultados promissores no desenvolvimento de modelos preditivos (Xiahou, Xiancheng e Harada, 2022). Porém, a aplicação dessas técnicas na previsão de *churn* em clubes de assinatura é pouco explorada na literatura.

Neste contexto, o objetivo desse trabalho foi comparar a performance de uma RNA *multilayer perceptron* (MLP) e um método composto por regressão logística e algoritmo média-K na previsão de *churn* utilizando dados de um clube de assinaturas que atua no mercado brasileiro. A definição do melhor método foi feita através da avaliação da medida-F1 e do retorno financeiro esperado para cada campanha de retenção.

Metodologia

Este trabalho seguiu a metodologia proposta por Bertrand e Fransoo (2002), que é apresentada resumidamente na Figura 1. Para a construção dos modelos foram utilizados dados da base de assinantes de um clube de assinaturas que atua no mercado brasileiro. As variáveis preditoras foram selecionadas e o banco de dados passou por um tratamento para remoção de dados faltantes e *outliers*. As variáveis selecionadas e seus respectivos valores possíveis são apresentados na Tabela 1.

Figura 1 – Método de pesquisa de modelagem e simulação.



Fonte: Adaptado de Bertrand e Fransoo (2002).

Tabela 1 – Variáveis preditoras utilizadas

Identificação	Notação	Descrição	Valores possíveis
fk_subscription	FK	Identificação numérica única de cada caso	
subscription_date	SD	Data de início da assinatura	De 28/07/2014 a 30/09/2023
product_detail	PD	Tipo de produto adquirido	0, 1 ou 2
payment_type	PT	Tipo de pagamento	Adiantado ou Recorrente
renewal_period	RP	Ciclo de renovação	Anual, Semestral, Trimestral ou Mensal
plan_price	PLP	Preço original do plano	Entre R\$ 20,00 e R\$ 90,00
paid_price	PDP	Preço pago atualmente pelo plano	Entre R\$ 20,00 e R\$ 90,00
subscription_cycle	SC	Parcela atual do plano	Entre 1 e 100
Desconto	D	Diferença entre preço pago e preço original	Entre R\$ 0,00 e R\$ 54,90

Fonte: Elaborado pelo próprio autor.

A Ciência do NANO e seu impacto transformador no MACRO

Tratamento do banco de dados: A base de dados original é composta por 170 mil assinantes, dos quais a taxa real de *churn* dentro da população analisada é de 8,9%, caracterizando assim um problema de desbalanceamento de classes (Chawla *et al.*, 2003). Para contornar esse problema, foi realizada uma subamostragem aleatória da classe majoritária (*random undersampling*) para a construção de uma amostra de 10.667 casos com uma taxa de *churn* de 25%. Essa amostra foi então utilizada para a construção dos modelos.

Método composto média-K e regressão logística: Utilizando o software Minitab 18®, foi aplicado à amostra o algoritmo média-K, que dividiu os dados em 2 subgrupos de acordo com a menor média entre os centroides de cada variável preditora. Cada subgrupo foi submetido à regressão logística, resultando dessa forma em duas equações de previsão que tinham como resultado a probabilidade de cancelamento de cada assinante. O valor limiar, estabelecido para separar os possíveis cancelamentos dos não-cancelamentos, foi de 30%.

Rede Neural MLP: uma RNA MLP composta por 7 neurônios na camada de entrada, 5 na camada oculta e 1 na camada de saída (MLP 7:7-5-1:1) foi construída utilizando o software Statistica 7®. Os dados da amostra foram separados em 3 grupos, sendo 50% deles utilizado para treinamento da rede MLP, 25% para validação e 25% para teste. A Tabela 2 mostra os parâmetros utilizados para construção da RNA MLP.

Tabela 2 – Parâmetros utilizados na construção da RNA MLP.

Grupo	Parâmetro	Descrição
Quantidade de neurônios	Camada de entrada	7
	Camada oculta	5
	Camada de saída	1
Parâmetros de treinamento	Algoritmo fase 1	Backpropagation
	Número de épocas	100
	Taxa de aprendizado	0,01
	Momentum	0,3
	Algoritmo fase 2	Gradiente descendente conjugado
	Número de épocas	500
Funções de ativação	Camada oculta	Tangente hiperbólica
	Camada de saída	Função logito

Fonte: Elaborado pelo próprio autor.

Análise de performance dos modelos: A qualidade de um modelo de previsão binário, como o de *churn*, pode ser medida tanto pela taxa de acertos do modelo entre as classificações positivas (precisão), quanto pela capacidade do modelo de encontrar positivos verdadeiros presentes na amostra (*recall*). Porém, analisar essas medidas isoladamente pode levar a conclusões equivocadas, principalmente quando uma das classes é significativamente maior que a outra (Kiguchi, Saeed e Medi, 2022). Dessa forma, utilizar uma relação entre a precisão e o *recall* (Medida F1) ou estimar o retorno financeiro esperado da aplicação de cada método ajudam na tomada de decisão em situações de aplicação real.

As Equações 1 a 4 (Lalwani *et al.* 2022 e Neslin *et al.* 2006) foram utilizadas para o cálculo dos indicadores de performance dos modelos, onde P_v são os positivos verdadeiros, P_f são os positivos falsos, N_v são os negativos verdadeiros, N_f são os negativos falsos, N é o número total de assinantes, η é a fração de assinantes impactados pela campanha de retenção, γ é a fração de cancelamentos reais que foram revertidos pela campanha, LTV é o valor adicional pago por cada cliente retido (*lifetime value*), d é o custo do incentivo oferecido para cada assinante, π_0 é o percentual de *churn* da amostra e λ é o coeficiente “*lift*” e mede o quão mais provável é o cancelamento de um assinante que foi contactado em relação ao total da amostra.

$$Recall = \frac{P_v}{P_v + N_f} \quad (1)$$

A Ciência do NANO e seu impacto transformador no MACRO

$$Precisão = \frac{P_v}{P_v + P_f} \quad (2)$$

$$Medida F1 = \frac{2 \times Precisão \times Recall}{(Precisão + Recall)} \quad (3)$$

$$Retorno = N \times \eta \times ((\gamma \times LTV + d \times (1 - \gamma)) \times \pi_0 \times \lambda - d) \quad (4)$$

A Tabela 3 apresenta as constantes utilizadas para cálculo do retorno financeiro das campanhas utilizando cada modelo investigado.

Tabela 3 – Constantes utilizadas para cálculo do retorno financeiro das campanhas.

Constante	Notação	Valor
Custo do incentivo	d	R\$ 30,00
Percentual de cancelamentos reais revertidos	γ	70%
Valor adicional pago por cada cliente retido	CLV	R\$ 150,00
Percentual de <i>churn</i> na amostra	π_0	25,0%

Fonte: Elaborado pelo próprio autor.

Resultados

A Tabela 4 apresenta o comparativo entre os indicadores de performance calculados para cada um dos modelos de previsão de *churn* avaliados, onde observa-se que o método composto por algoritmo média-K e regressão logística (MK+RL) apresentou uma precisão de 55,9%, sendo superior aos 51,7% atingidos pela RNA MLP, demonstrando uma maior proporção de positivos verdadeiros dentre todos os positivos classificados pelo modelo. Porém a RNA MLP apresentou todos os demais indicadores superiores ao método MK+RL, com um *recall* de 38,3% versus 24,7% da MK+RL, o que demonstra uma maior proporção de positivos verdadeiros dentre todos os positivos disponíveis na amostra, considerando tanto os positivos verdadeiros (P_v) quanto os negativos falsos (N_f), ou seja, mais cancelamentos reais são impactados por campanhas construídas a partir deste modelo.

Tabela 4 – Indicadores de performance dos modelos avaliados.

Indicador	MK+RL	Rede MLP
Recall	24,7%	38,3%
Precisão	55,9%	51,7%
Medida – F1	34,2%	44,0%
Retorno da Campanha (R\$)	39.702	57.258

RL: Regressão logística; MK – Algoritmo média-K; MLP: *Multilayer perceptron*.

Fonte: Elaborado pelo próprio autor.

Quando se analisa o impacto desses indicadores na performance final dos modelos, a RNA MLP apresentou uma medida – F1 de 44,0% versus 34,2% do método composto. O mesmo comportamento é observado quando analisado o retorno esperado sobre a campanha de retenção, onde a rede MLP apresentou um lucro 44,2% maior do que o método MK+RL, o que do ponto de vista de mercado é um indicador significativamente expressivo.

Discussão

A Ciência do NANO e seu impacto transformador no MACRO

Quando analisado de forma ampla, observa-se que a RNA MLP apresentou uma performance de previsão superior ao método composto pelo algoritmo média-K e regressão logística. Principalmente considerando que o objetivo principal da construção de um modelo de previsão de *churn* é construir campanhas de retenção que tenham o maior impacto positivo possível na lucratividade das empresas (Li e Li, 2019). Isso pode ser explicado pelo maior *recall* que pode ser traduzido em uma maior captura de positivos verdadeiros pela rede MLP, que por sua vez torna possível uma maior recuperação das campanhas de retenção regidas por esse modelo.

A previsão de *churn* em clubes de assinatura utilizando RNA MLP é pouco explorada na literatura. Hoang e Cam (2025) utilizaram redes neurais para prever o *churn* de uma empresa do mercado de jogos eletrônicos e atingiram uma precisão de 66,7%. AbdelAziz et al. (2025) afirmam que a combinação de redes neurais com técnicas do *clustering* ou algoritmos de *boosting* como o XGboost pode contribuir para melhorar a performance de modelos de previsão de *churn*, o que poderia ser uma estratégia de continuidade e melhoria para o modelo atual proposto.

Da mesma forma, Xiahou e Harada (2022) aplicaram uma combinação de média-K e regressão logística para previsão de *churn* no mercado de *e-commerce* atingindo uma precisão de 90,7%. Os autores otimizaram o modelo ao definir um número ótimo de subgrupos gerados pelo algoritmo média-K em 3, aplicando regressão logística separadamente à cada subgrupo, tal estratégia pode trazer mais fidelidade e robustez a modelos de previsão regressivos.

No contexto de mercado onde se encontra o clube de assinaturas objeto de estudo desse trabalho, o uso de redes neurais artificiais mostrou ter potencial para ajudar a diferenciar o perfil de assinantes com maior e menor probabilidade de cancelamento e, a partir disso, orientar a construção de campanhas de retenção que tenham impacto significativamente positivo na rentabilidade do negócio.

Conclusão

Dentro do contexto de negócio do mercado de compras recorrentes, a aplicação de ferramentas de previsão de *churn* pode se tornar um importante diferencial estratégico das empresas deste ramo, já que permite uma compreensão mais ampla de quais são os fatores que definem a decisão de cancelamento de um assinante. A capacidade de simulação de relações não lineares entre as variáveis preditoras faz com que as RNA se destaquem como uma ferramenta poderosa e capaz de processar grandes quantidades de informação de forma eficiente e gerar previsões acuradas, se mostrando uma ferramenta superior ao modelo composto por média K com regressão logística estudado, trazendo um retorno financeiro superior no cenário explorado, permitindo traçar estratégias que, aplicadas à construção de campanhas de retenção, vão reduzir a taxa de cancelamentos, aumentando o faturamento da empresa e a percepção de valor por parte do cliente.

Referências

- ABDELAZIZ, N. M.; BEKHEET, M.; SALAH, A.; EL-SABER, N.; ABDELMONEIM, W. T. A Comprehensive Evaluation of Machine Learning and Deep Learning Models for Churn Prediction. **Information**, v. 16, n. 7, p. 537, 25 jun. 2025.
- BAHNSEN, A. C.; AQUADA, D.; OTTERSTEN, B. A novel cost-sensitive framework for customer churn predictive modeling. **Decision Analytics**, v. 2, n. 1, dez. 2015.
- BIANCHESI, N. M. P.; MATTA, C. E. DA; STREITENBERGER, S. C.; ROMÃO, E. L.; BALESTRASSI, P. P.; COSTA, A. F. B. A nonlinear time-series prediction methodology based on neural networks and tracking signals. **Production**, v. 32, 2022.
- CAIGNY, A. DE; COUSSEMENT, K.; BOCK, K. W. DE. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. **European Journal of Operational Research**, v. 269, n. 2, p. 760–772, set. 2018.
- CHAWLA, N. V.; LAZAREVIC, A.; HALL, L. O.; BOWYER, K. W. SMOTEBoost: Improving Prediction of the Minority Class in Boosting. Em: **Knowledge discovery in databases> PKDD 2003**. [s.l.] Springer, 2003. v. 2838p. 107–119.
- CHEN, T.; FENYO, K.; YANG, S.; ZHANG, J. **Thinking inside the subscription box: new research on e-commerce consumers**. Disponível em: <<https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/thinking-inside-the-subscription-box-new-research-on-e-commerce-consumers>>. Acesso em: 23 maio. 2022.

A Ciência do NANO e seu impacto transformador no MACRO

- HOANG, H. D.; CAM, N. T. Do they like your game? Early-stage churn prediction using a two-phase neural network system. **Engineering Applications of Artificial Intelligence**, v. 144, p. 110102, mar. 2025.
- IMARC. **Subscription Box Market Report by Type (Replenishment Subscription, Curation Subscription, Access Subscription), Gender (Male, Female), Application (Clothing and Fashion, Beauty, Food and Beverages, Pet Food, Baby Products, Health and Fitness, and Others), and Region 2024-2032**. [s.l.: s.n.]. Disponível em: <<https://www.imarcgroup.com/subscription-box-market>>. Acesso em: 31 out. 2023.
- KIGUCHI, M.; SAEED, W.; MEDI, I. Churn prediction in digital game-based learning using data mining techniques: Logistic regression, decision tree, and random forest. **Applied Soft Computing**, v. 118, 2022.
- LI, X.; LI, Z. A hybrid prediction model for e-commerce customer churn based on logistic regression and extreme gradient boosting algorithm. **Ingenierie des Systemes d'Information**, v. 24, n. 5, p. 525–530, 2019.
- SEETHARAMAN, P. Business models shifts: Impact of Covid-19. **International Journal of Information Management**, v. 54, 1 out. 2020.
- WILL M. BERTRAND, J.; FRANSOO, J. C. Operations management research methodologies using quantitative modeling. **International Journal of Operations & Production Management**, v. 22, n. 2, p. 241–264, fev. 2002.
- XIAHOU, X; HARADA, Y. B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM. **Journal of Theoretical and Applied Electronic Commerce Research**, v. 17, n. 2, p. 458–475, 2022.
- XIAHOU, XIANCHENG; HARADA, Y. B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM. **Journal of Theoretical and Applied Electronic Commerce Research**, v. 17, n. 2, p. 458–475, 6 abr. 2022.
- ZHANG, T.; MORO, S.; RAMOS, R. F. A Data-Driven Approach to Improve Customer Churn Prediction Based on Telecom Customer Segmentation. **Future Internet**, v. 14, n. 3, 2022.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).