

MODELAGEM ESTATÍSTICA DA INADIMPLÊNCIA SOB A PERSPECTIVA DA CIÊNCIA DA INFORMAÇÃO

STATISTICAL MODELING OF DEFAULT FROM AN INFORMATION SCIENCE PERSPECTIVE

Juliana Simoes da Silva Dutra

Universidade Federal de Minas Gerais, UFMG

E-mail: jussdutra@gmail.com

ORCID: 0009-0003-1663-899X

Frederico Cesar Mafra Pereira

Universidade Federal de Minas Gerais, UFMG

E-mail: professorfredericomafra@gmail.com

ORCID: 0000-0002-1971-8069

Currículo Lattes: 4271090306897767

RESUMO

O crescimento da inadimplência representa desafio para instituições financeiras e empresas de recuperação de crédito, exigindo estratégias eficientes de análise, segmentação e tomada de decisão. A Ciência da Informação, associada à Estatística e às tecnologias de análise de dados, contribui para a transformação de grandes volumes de dados em conhecimento aplicado à gestão organizacional. O estudo apresenta uma análise da inadimplência sob a perspectiva da Ciência da Informação, utilizando técnicas estatísticas para construção de modelo preditivo aplicado à recuperação de crédito. O objetivo foi identificar variáveis relacionadas à probabilidade de inadimplência de clientes inadimplentes, além de discutir o papel da gestão da informação e inteligência analítica na construção de estratégias de cobrança mais eficientes. 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, cedidas por um escritório de cobrança extrajudicial de Belo Horizonte/MG. O modelo estatístico adotado foi a regressão logística, com seleção de variáveis por stepwise forward selection e avaliação por Curva ROC e AUC. 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. O modelo apresentou AUC aproximada de 0,70,

indicando capacidade preditiva moderada. Conclui-se 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: Ciência da Informação, Inadimplência, Regressão Logística, Gestão da Informação, Inteligência Artificial.

ABSTRACT

The growth of delinquency represents a challenge for financial institutions and debt recovery companies, demanding efficient strategies for analysis, segmentation, and decision-making. Information Science, associated with Statistics and data analysis technologies, contributes to the transformation of large volumes of data into knowledge applied to organizational management. This study presents an analysis of delinquency from the perspective of Information Science, using statistical techniques to build a predictive model applied to debt recovery. The objective was to identify variables related to the probability of repayment among defaulting clients, as well as to discuss the role of information management and analytical intelligence in building more efficient collection strategies. The research is characterized as quantitative, descriptive, and exploratory. The database, anonymized and in compliance with the LGPD (Brazilian General Data Protection Law), comprised information from 665,379 delinquent credit card clients, provided by an extrajudicial debt collection office in Belo Horizonte/MG. The statistical model adopted was logistic regression, with variable selection by stepwise forward selection and evaluation by ROC curve and AUC. The results showed that 96.2% of clients were classified as bad payers and 3.8% as good payers. Marital status, debt phase, debt amount, age, gender, and due date were the most relevant variables. The model presented an approximate AUC of 0.70, indicating moderate predictive capacity. It is concluded that the integration between Information Science, Statistics, and Data Science favors more efficient, segmented collection strategies aligned with organizational knowledge management.

Keywords: Information Science, Default, Logistic Regression, Information Management, Artificial Intelligence.

REFERÊNCIAS

BANCO CENTRAL DO BRASIL. Resolução CMN nº 4.966, de 25 de novembro de 2021.

CHOO, C. W. **A organização do conhecimento**. São Paulo: Senac, 2003.

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

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

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

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