

Análise de Risco de Crédito para Gestoras de Vendedores Porta-a-Porta usando Aprendizado de Máquina^{*}

Ricardo Zorzal Davila^{*,**} Francisco de Assis Boldt^{*} Filipe Mutz^{*}

^{*} Programa de Pós-Graduação em Computação Aplicada, Instituto Federal do Espírito Santo (IFES) - Campus Serra, ES
(emails: {franciscoa, filipe.mutz}@ifes.edu.br).

^{**} Máximus Soluções, ES (e-mail: ricardozorzaldavila@gmail.com).

Abstract: Credit risk analysis is fundamental for small and medium companies given that defaults cause significant impacts in revenue. Banks and other financial institutions have access to several information regarding the financial wellbeing of clients for credit analysis. On the other hand, companies from other areas need to estimate credit risk using demographic information and the history of interactions with the company. This work studies the problem of credit risk analysis for companies that manage door-to-door salesmen. These companies have specific characteristics that differentiate them from businesses of other niches. These specificities influence the process of credit risk analysis. A dataset was built using information from partner companies and this dataset was used to train several machine learning algorithms in the task of default prediction. Experiments showed that the logistic regression classifiers achieves the same performance as nonlinear classifiers (e.g., neural networks, XGBoost, and ensembles), while being less prone to overfitting and being more interpretable.

Resumo: A análise de risco de crédito é de fundamental importância para pequenas e médias empresas uma vez que o não pagamento de compras gera impactos significativos nos rendimentos. Bancos e financeiras possuem acesso à diversas informações relacionadas à saúde financeira dos clientes para análise de crédito. Já empresas de outros ramos precisam calcular o risco de crédito usando informações demográficas de clientes e seu histórico de interações com a empresa. Este trabalho estuda o problema de análise de risco de crédito para empresas gestoras de vendedores porta-a-porta. Estas empresas possuem características específicas que as diferenciam de negócios de outros segmentos. Estas especificidades influenciam o processo de análise de risco de crédito. Foi construída uma base de dados usando informações de empresas parceiras e esta base foi usada para treinar diversos algoritmos de aprendizado de máquina na tarefa de predição de não pagamento. Experimentos mostraram que o classificador linear *logistic regression* é capaz de alcançar a mesma performance de classificadores não lineares (e.g., redes neurais, XGBoost e *ensembles*), sendo menos suscetível à *overfitting* e mais interpretável.

Keywords: Credit Risk Analysis; Credit Scoring; Artificial Intelligence; Machine Learning; XGBoost; Neural Networks; Logistic Regression.

Palavras-chaves: Análise de Risco de Crédito; Pontuação de Crédito; Inteligência Artificial; Aprendizado de Máquina; XGBoost; Redes Neurais; *Logistic Regression*.

1. INTRODUÇÃO

O setor de crédito é de considerável importância econômica. Em 2019, o crédito ao consumidor em aberto nos EUA era de aproximadamente US\$ 4.11 trilhões (Federal Reserve System, 2021; Statista, 2021). No Brasil, as estatísticas são similares. Em fevereiro de 2021, o saldo total

de empréstimos e financiamentos ao setor não financeiro no Brasil era de aproximadamente R\$ 4.3 trilhões (Banco Central do Brasil, 2021).

O número de empréstimos e a taxa de inadimplência no pagamento de empréstimos é influenciada pela situação social, política e econômica dos países. De 2011 a 2021, a taxa de inadimplência de carteiras de crédito para pessoas físicas oscilou entre 3% e 4,6%, alcançando valor máximo de 5,5% no Brasil. Mesmo com a política de juros em caso de atraso de pagamento, empresas que optam por dar crédito aos clientes podem sofrer impactos financeiros devido à inadimplência. Estes impactos são particularmente significativos em empresas de pequeno e médio porte. Buscando minimizar estes impactos, é comum o uso de

^{*} Agradecemos à FAPES e a CAPES pelo apoio financeiro dado por meio do PDGP (Parcerias Estratégicas nos Estados da CAPES) (PROCESSO: 2021-2S6CD, TO/nº FAPES: 132/2021). Também agradecemos ao Propós (Programa Institucional de Apoio à Pós-graduação Stricto Sensu) do IFES pela apoio financeiro. Filipe Mutz agradece ao Instituto Federal do Espírito Santo (IFES) por incentivar sua pesquisa via o Programa Pesquisador de Produtividade (PPP) - portaria n. 1072 de 21 de maio de 2020.