A APLICAÇÃO DE APRENDIZADO DE MÁQUINA PARA MODELOS BINÁRIOS DE ANÁLISE DE CRÉDITO

Raffael Guideti Miello

Centro Universitário UNICESUMAR

Março de 2024

**Resumo:**

Neste trabalho, foi abordado diversas técnicas de análise de dados por meio de bibliotecas na linguagem de programação python no intuito de aplicá-las a machine learning (aprendizado de máquina) para a melhor estruturação de dados e compreender por meio de modelo matemáticos binário quais consumidores do serviço de crédito são “maus ou bons pagadores”.

Foi integrada a biblioteca PowerSHAP, que permite uma análise interpretável do modelo por meio do método SHAP (SHapley Additive exPlanations). Essa abordagem possibilitou entender a influência de cada variável na decisão do modelo, garantindo maior transparência e explicabilidade dos resultados. O pipeline estruturado seguiu uma arquitetura modular, organizando as funções de logging, carregamento de dados, modelagem preditiva e análise interpretável em diferentes camadas para facilitar a manutenção e reuso do código.

Dessa forma, este trabalho apresenta uma solução completa para análise preditiva de risco de crédito, combinando técnicas de engenharia de dados, aprendizado de máquina e interoperabilidade de modelos, contribuindo para a tomada de decisão no setor financeiro.

**Palavras-chave:** Análise de crédito, Machine learning, PowerSHAP.

**1. INTRODUÇÃO:**

A análise de risco de crédito é um dos desafios fundamentais no setor financeiro, influenciando diretamente a concessão de empréstimos e a gestão de inadimplência. Com o avanço do aprendizado de máquina, tornou-se possível desenvolver modelos preditivos capazes de classificar consumidores com base em seu histórico e comportamento financeiro. Neste trabalho, propõe-se a construção de um pipeline de Machine Learning utilizando a linguagem Python para prever a probabilidade de um solicitante ser um “bom” ou “mau pagador”. A abordagem inclui técnicas de pré-processamento de dados, modelagem com algoritmos de classificação e explicabilidade dos resultados por meio do método SHAP, garantindo maior transparência e interoperabilidade ao modelo.