



Modelo Preditivo de Inadimplência em Empréstimos P2P utilizando Algoritmos de Classificação Supervisionada

Theodoro Priosti de Almeida¹

¹Faculdade de Computação e Informática (FCI)
Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

theodoro.almeida@mackenzie.br

Resumo. Este projeto visa desenvolver um modelo de análise preditiva para avaliação de risco de crédito utilizando o dataset Lending Club, uma das maiores plataformas de empréstimos peer-to-peer dos Estados Unidos.

Através de machine learning, pretende-se criar um sistema capaz de classificar solicitações de empréstimo quanto ao risco de inadimplência, contribuindo para decisões mais assertivas no mercado financeiro.

1. Introdução

O mercado de crédito desempenha um papel fundamental na economia global, facilitando o acesso ao capital para indivíduos e empresas. Com o advento das fintechs e plataformas de empréstimos, surgiram novas oportunidades e desafios na avaliação de risco de crédito. A Lending Club, fundada em 2006, tornou-se pioneira nesse segmento, conectando investidores a tomadores de empréstimo.

A avaliação tradicional de risco de crédito baseava-se principalmente em scores como o FICO e informações básicas dos solicitantes. No entanto, a disponibilidade de grandes volumes de dados e o desenvolvimento de algoritmos proporcionaram novas possibilidades para análises mais sofisticadas e precisas. A necessidade de modelos preditivos mais eficazes para avaliação de risco de crédito é evidente considerando:

- Crescimento do mercado: O setor de empréstimos digitais movimenta bilhões de dólares anualmente, demandando ferramentas de análise cada vez mais precisas.
- Redução de perdas: Modelos preditivos podem significativamente reduzir as taxas de inadimplência, protegendo investidores e instituições financeiras.
- Democratização do crédito: Algoritmos sofisticados podem identificar bons pagadores que seriam rejeitados por métodos tradicionais.

O objetivo final é desenvolver um modelo de machine learning para predição de risco de inadimplência em empréstimos, utilizando dados históricos da Lending Club, visando melhorar a precisão da análise de crédito e reduzir perdas financeiras.

O projeto se enquadra na opção “Framework”, onde será usado a linguagem Python com bibliotecas de Machine Learning para chegar no objetivo final.

2. Descrição do problema

O problema central abordado neste projeto é a predição de inadimplência em empréstimos pessoais, um desafio crítico no setor financeiro que envolve múltiplas dimensões. Naturalmente, a maioria dos empréstimos são pagos, criando um desafio de classificação com classes desbalanceadas

- Existem centenas de constantes que podem influenciar o risco de crédito
- Aprovar um empréstimo que entrará em default resulta em perda direta
- Rejeitar um bom pagador representa perda de oportunidade de lucro

3. Aspectos Éticos do Uso da IA

O desenvolvimento de sistemas de IA para análise de crédito levanta importantes questões éticas e de responsabilidade que devem ser cuidadosamente consideradas.

Será necessário garantir que o algoritmo não discrimine contra nenhum grupo e será necessário atualização e monitoramento constante para garantir futura precisão.

Dados pessoais dos usuários dever ser corretamente anonimizados, e os usuários clientes do uso.

4. Dataset

O dataset utilizado provém da Lending Club, uma das maiores plataformas de empréstimos dos Estados Unidos. Os dados foram disponibilizados publicamente pela empresa e cobrem o período de 2007 a 2018, representando um dos mais abrangentes conjuntos de dados sobre empréstimos disponíveis.

O dataset está disponível em [Kaggle \[wordsforthewise/lending-club\]](#), e foi pré processado pelo autor. O dataset está licenciado sob a licença CC0: Domínio Público.

A análise exploratória, [disponível no Github](#), revelou que a maioria dos empréstimos são pagos integralmente, com uma distribuição específica entre diferentes status de inadimplência, confirmando o desafio de classes desbalanceadas. O volume de empréstimos aumentou significativamente nos anos analisados, sendo a finalidade principal a consolidação de dívidas.

5. Metodologia e Resultados

O projeto usará machine learning para prever inadimplência. Serão testados diversos algoritmos e avaliados por precisão.

Esperamos reduzir 20% as taxas de inadimplência e automatizar decisões de baixo risco, enquanto tornamos os critérios mais transparentes.

6. Referências

Github: github.com/ovosimpatico/predictive-credit

Dataset: kaggle.com/datasets/wordsforthewise/lending-club

7. Bibliografia

Khandani, Amir & Kim, Adlar & Lo, Andrew. (2010). Consumer Credit-Risk Models Via Machine-Learning Algorithms. *Journal of Banking & Finance*. 34. 2767–2787. 10.1016/j.jbankfin.2010.06.001.

Lessmann, Stefan & Baesens, Bart & Seow, Hsin-Vonn & Thomas, Lyn. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*. (doi:10.1016/j.ejor.2015.05.030). 10.1016/j.ejor.2015.05.030.