

APRP - Projeto Final

Duração: 2 semanas

Entrega:

- Entrega final: **30/05/2023**

Descrição:

O projeto tem como principal objetivo permitir estender os conhecimentos adquiridos na componente Teórica e Prática da disciplina. O trabalho pode ser desenvolvido individualmente ou em grupo de 2 elementos.

Contexto:

As empresas de empréstimos P2P (*peer-to-peer*) tiveram uma forte expansão na última década. Através do recurso a plataformas web estas empresas permitem conectar credores e devedores. Para o devedor é proposto um processo de aprovação menos burocrático, tendo o credor como principal motivação obter um retorno mais elevado.

Na medida em que muitos destes empréstimos não possuem colateral um dos desafios consiste na avaliação do nível de risco associado a estes depósitos. O dataset a usar contém os dados demográficos e financeiros de todos os empréstimos de uma empresa P2P entre fevereiro de 2019 e julho de 2021.

É objetivo deste projeto focar na previsão do *Status* que pode adquirir os valores: *Late*, *Repaid*, *Current*. É ainda objetivo aplicar uma *Bayes net* que possibilite a interpretação dos dados.

O dataset tem um total de 112 *features* relativas a cerca de 180k empréstimos. Uma descrição detalhada, assim como o download do csv pode ser realizada em:

<https://www.kaggle.com/datasets/sid321axn/bondora-peer-to-peer-lending-loan-data>

Plano de trabalho:

1. Obtenção do dataset e tratamento adequado dos dados raw. (*Data Cleaning, missing data; Data discretization; ...*).
2. Análise Exploratória. Deve verificar no Kaggle outros Notebooks já realizados sobre este dataset. Confirmar se necessário tratamento adicional.
3. Implementação de uma estratégia de classificação baseada em *Naive Bayes* que deve ser utilizada como baseline.
4. **Aprendizagem dos parâmetros da *Bayes net*.**
5. **Análise da *Bayes net* obtida.**
6. **Demonstração da utilização da rede para obtenção de inferências.**
7. Preparação de relatório que descreva o trabalho realizado, os resultados obtidos assim como uma análise crítica dos mesmos.
8. Deve ser entregue o relatório no formato PDF, assim todo o código usado preferencialmente em Python e Jupyter Notebook.