

Engenharia de Software para Ciência de Dados - PUC-Rio

Trabalho Final

Prof. Marcos Kalinowski

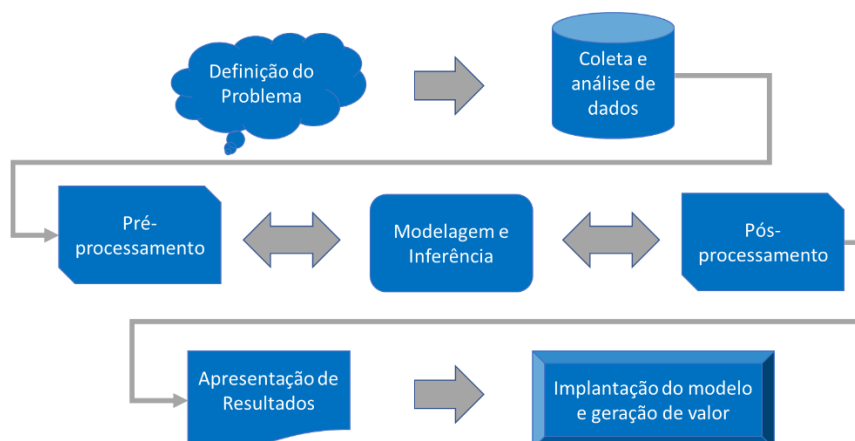
Assunto

Comparação de Modelos de Machine Learning de Classificação (ou Regressão) Utilizando Boas Práticas de Engenharia de Software

Cenário

Este trabalho deverá ser feito individualmente e deverá ser entregue até às 23h59 do dia determinado.

Escolha uma base de dados para um problema de classificação ou regressão. Sugere-se usar uma das bases de dados disponibilizada no UCI Machine Learning Repository ou no Kaggle. Você deverá trabalhar desde o problema até os resultados e necessariamente comparar e otimizar os hiperparâmetros para mais de um modelo.



Lembre-se das etapas para implementação de projetos de ciência de dados, apresentadas em sala de aula, e trabalhe com esta base de dados no ambiente Google Colab, utilizando Python e a biblioteca Scikit-Learn. Produza um notebook que servirá como relatório, descrevendo textualmente cada uma das etapas do seu código. Para a definição do problema, descreva uma user story e seu detalhamento (simplificado – não é necessário utilizar o template de especificação da atividade anterior, já que é um *toy problem* e não um projeto industrial), na primeira célula do notebook. Para a implementação, sugere-se utilizar as boas práticas de programação orientada a objetos trabalhadas em sala.

Na última célula, informe como você implantaria a solução de forma a disponibilizar o resultado do seu modelo para o usuário final.