

Aprendizagem de máquina

Atividade avaliativa – 01 – Porto Seguro Data Challenge

Especificação

Nesta atividade você devem formar duplas para participar de um desafio proposto pelo Porto Seguro. O desafio está disponível na plataforma do Kaggle no link "<https://www.kaggle.com/competitions/porto-seguro-data-challenge/overview>". Na página da competição tem a descrição do problema, informações da base dados e exemplos do arquivo de submissão.

Tópicos para considerar

- Levar em consideração os tipos das colunas (quantitativo e qualitativo)
- Imputação de valores faltantes (missing value imputation)
- Base de dados desbalanceada (resampling: oversampling e undersampling)
- Métodos de validação do modelo (cross validation)
- Redução de dimensionalidade (dimensionality reduction)
- Seleção de atributos (features selection)
- Engenharia de atributos (feature engineer)
- Utilizar e comparar as diferentes técnicas de classificação disponíveis na biblioteca "scikit learn" (Logistic Regression, Suport Vector Machine, Naive Bayes, KNN, Decision Tree, etc.)

Entregáveis

- Relatório explicando os principais pontos da solução (.pdf ou no próprio notebook).
- Notebook desenvolvido em python.

Realize a submissão pelo Teams.

Obs.: apenas um integrante da dupla precisa realizar a submissão, mas especifique os nomes no relatório. Prazo: 30/04