

**Projeto de Software Orientado a Dados**  
**Trabalho Prático 02 – Engenharia de Atributos e Ajustes de Modelos**

**Professor:** Fabrício A. Silva  
**Data de Entrega:** 22/02/2022  
**Valor:** 10 pontos  
**Trabalho individual**

## **Objetivo**

O objetivo deste trabalho é aplicar técnicas de engenharia de atributos e ajustar os parâmetros do algoritmo adotado para melhorar os resultados de classificação do problema de previsão do conjunto de dados Titanic.

## **Atividades**

1. Após entender e conhecer bem os atributos do conjunto de dados, aplicar técnicas de engenharia de atributos em dados numéricos e categóricos. A decisão deve ser baseada em fundamentos de acordo com o problema e as características dos dados.

2. Após a etapa 1, avaliar a necessidade e aplicar técnicas de seleção de atributos, para que sejam utilizados apenas os atributos mais relevantes para o modelo.

3. Estudar os detalhes do algoritmo a ser utilizado. Como ele funciona? Quais são seus parâmetros e o que representam? Quais as vantagens e desvantagens? Existe alguma restrição? Em quais tipos de problemas ele funciona melhor? Preparar uma apresentação de 10min sobre o algoritmo e fazer um vídeo com a apresentação.

**Lista de algoritmos:** Logistic Regression, Perceptron Algorithm, Naive Bayes, Decision Trees, Random Forests, AdaBoost, Neural Network, SVM, XGBoost.

**Escolha:** a escolha será feita junto com o professor, para que cada aluno fique com um algoritmo distinto.

4. Aplicar o algoritmo escolhido no conjunto de dados já com os atributos preparados após a etapa de engenharia de atributos. Ajustar os parâmetros do algoritmo para melhorar o seu desempenho. Aplicar validação cruzada (com dados de treino, validação e teste) para avaliar com os dados disponíveis para treino.

5. Revisitar os passos 1, 2 e 4 quantas vezes forem necessárias. Após chegar ao melhor cenário, submeter os resultados para o Kaggle e comparar os resultados com a primeira submissão feita. Os resultados melhoraram significativamente?

6. Aplicar alguma técnica de Auto-ML nos dados e comparar os resultados com os obtidos pelo seu modelo.

## **Forma de Entrega**

Entregar via Moodle um relatório contendo todo o código utilizado, os slides e descrevendo todas as etapas e decisões tomadas. Disponibilizar o link do vídeo no Youtube. Apresentar os resultados da validação cruzada para as várias configurações testadas. Este trabalho é **individual**.