

# Aprendizagem Automática II

## Proposta para o Trabalho Prático

-

Mestrado em Engenharia Informática  
Universidade do Minho

### Grupo

---

PG41080	João Ribeiro Imperadeiro
PG41081	José Alberto Martins Boticas
PG41091	Nelson José Dias Teixeira
PG41851	Rui Miguel da Costa Meira

4 de Abril de 2020

## 1 Introdução

Tal como foi requerido pelo docente desta disciplina, este documento serve para fazer um levantamento preliminar do que será desenvolvido para o respetivo trabalho prático. Exibe-se de seguida o repositório *GitHub* público associado a este projeto onde será colocado todo o código desenvolvido pelo grupo bem como toda a documentação e recursos intrínsecos à implementação:

`https://github.com/Nelson198/AA2`

## 2 Tema

O tema escolhido pelo nosso grupo é o desenvolvimento de uma *framework* de AutoML. A *framework* visa obter o melhor modelo para problemas de *supervised learning* e *unsupervised learning*. O objetivo final é colocar a *framework* disponível para os utilizadores de Python.

## 3 Planificação

Devido à complexidade do desenvolvimento deste projeto num prazo tão curto não estamos a incluir o pré processamento de dados neste projeto. A *framework* recebe como parametros os dados de treino e os dados de teste. A *framework* vai distinguir entre um problema de *supervised* e *unsupervised* consoante receba as *labels/targets* ou não. Para problemas de *supervised* vai ser distinguido entre regressão e classificação se o *target* é uma variavel continua ou discreta. Em problemas de regressão os algoritmos usados para encontrar o melhor modelo são: Regressão Linear, Regressão Polinomial, Support Vector

Regression, Decision Tree Regression, Random Forest Regression e Redes Neurais. Para problemas de classificação os algoritmos usados para encontrar o melhor modelo são: Regressão Logística, KNN, Support Vector Machine, Kernel SVM, Naive Bayes, Decision Tree Classification, Random Forest e Redes Neurais. Para problemas de clustering os algoritmos em uso para procurar o melhor modelo são: k-Means Clustering e Hierarchical Clustering. A procura de melhores *hyperparameters* para Redes Neurais vai ser realizada com o uso da biblioteca **kerastuner**. Para os outros algoritmos vai-se fazer uso do GridSearchCV e RandomizedSearchCV para a escolha dos melhores *hyperparameters*.

## 4 Objetivos

1. Garantir que ...
2. Verificar que