# Trabalho Prático: Aprendizagem Automática I – Parte 1

**Elementos:**

* Carlos José Lima Gonçalves, a77278
* José Pedro dos Santos Ferreira, a78452
* Ricardo Jorge Marques Peixoto, a78587

***Dataset*:** diabetes

O problema consiste em prever se um determinado indivíduo contém ou não diabetes, consoante alguns fatores.

Do conjunto de dados fazem parte as seguintes variáveis:

* Id: ID da pessoa (não é para considerar)
* Chol: colesterol total
* stab.glu: glucose estabilizada
* hdl: Lipoproteína de alta densidade
* ratio: rácio entre colesterol/hdl
* glyhb: hemoglobina glicada
* location: país (um fator com opções Buckingham e Louisa)
* age: idade em anos
* gender: género (fator com opções masculino e feminino)
* height: altura em polegadas
* weight: peso em libras
* frame: um fator com opções pequeno, médio e grande
* bp.1s: 1º pressão arterial sistólica (quando o coração bombeia sangue)
* bp.1d: 1º pressão arterial diastólica (quando o coração não está a bombear o sangue)
* bp.2s: 1º pressão arterial sistólica
* bp.2d: 1º pressão arterial sistólica
* waist: cintura em polegadas
* hip: anca em polegadas
* time.ppn: tempo, em minutos, após a refeição

A variável resposta depende do nível de hemoglobina glicada, visto que o seu valor acima de 7 é geralmente considerado um teste positivo de diabetes.

Ao observarmos estes dados vemos que se trata de um problema supervisionado, de classificação.

Podemos realizar este trabalho utilizando técnicas de classificação nas quais obtemos a probabilidade de ter diabetes, no entanto e visto que conhecemos o método de classificação de ter ou não diabetes, podemos também usar um método de XXXXX para prever o valor da hemoglobina e depois classificar consoante o valor obtido.

**Questões às quais vamos tentar responder:**

* será que o colesterol ou pressão arterial afetam os diabetes?
* qual a probabilidade (ou confiança no resultado) de uma pessoa com o colesterol X ter diabetes?