

# Trabalho 1 de Aprendizagem Automática

João Cavaco  
nº 142470@alunos.uevora.pt

1 de fevereiro de 2021



## Resumo

O objectivo deste trabalho é implementar em Python 3 uma classe para gerar árvores de decisão com “pruning”

## 1 Organização do código

O código está organizado da seguinte forma:

- Funções relacionadas com a pureza e classificação dos dados:
  - Information\_gain
  - Classification\_error
  - Gini
  - Entropy
  - Homogeneous
  - Label
- Funções relacionadas com a partição do conjunto de dados:
  - determine\_best\_split
  - split\_data
- Classe Node que representa um nó da árvore de decisão com os seguintes atributos.
  - prediction : tuplo que guarda o id da coluna e o valor em é feita a divisão de dados nesse nó da árvore que caso o atributo leaf seja True é interpretado como a classe a ser atribuída aos valores desse nó.
  - left
  - right
  - leaf : indica se o nó é uma folha.
- Classe DecisionTree uma árvore de decisão e as suas funções:
  - fit
  - growTree
  - predict
  - score

O código encontra-se comentado para que seja mais fácil entender o funcionamento de cada função mas o funcionamento geral do programa encontra-se na secção seguinte.

## 2 Funcionamento do programa

1. Os dados do ficheiro csv são importados.
2. Os dados são separados em dados de teste e dados de treino
3. É criada uma árvore de decisão com uma função de pureza à escolha.
4. Os dados de treino são fitted à árvore de decisão.
  - 4.1. É chamada a função recursiva growTree que funciona de acordo com o algoritmo ID3.
5. A árvore classifica um uma nova instância de um conjunto de dados.
  - 5.1. É chamada a função recursiva predict que percorre a árvore comparando os valores em cada nó até chegar a uma folha.
6. Para calcular a precisão da classificação do modelo é efetuada a classificação do conjunto de dados de teste e depois as etiquetas previstas são comparadas com as etiquetas reais.

## 3 Limitações do programa

Não foi implementado nenhum método de poda da árvore.

## 4 Análise de desempenho

Conjunto de dados	Gini	Entropia	Error
vote	93,10%	93,10%	58.62%
contact-lenses	100%	100%	83.33%
soybean	91,49%	91,49%	79,43%
weather.nominal	50,00%	50,00%	25,00%