

# Classificação dos alunos da disciplina Aprendizado de Máquina usando Decision Tree

Aprovados ou Reprovados

Diogo Akio Balboni Miyake  
Luiz Gabriel Correia  
Marcelo de Souza Pena  
Thais Reis Alves

# Atributos

- Notas (AED1, PI, IA, MD e Algelin)
- Número de disciplinas trancadas
  - Número de conceitos O
    - Número de faltas
      - CR

# Base de dados

- Gerada pelo grupo
  - 70% aprovados e 30% reprovados
- Horas de estudo: normais com centro em 4,0 para aprovados e em 1,5 para reprovados
- Notas: de 0 a 10 com base nas horas de estudo

# Base de dados

- Disciplinas trancadas: aleatoriamente entre 0 e 4 para aprovados e entre 2 e 6 para reprovados;
- Conceitos O: aleatoriamente entre 0 e 1 para aprovados e entre 0 e 4 para reprovados;
- Faltas: entre 0 e 2 para aprovados e entre 0 e 6 para reprovados;

# Base de dados

- CR: normal centrado em 2,4 com desvio de 0,5 para aprovados e centrado em 1,6 com desvio de 0,5 para reprovados

# Base de dados

Exemplo:

|                       |          |
|-----------------------|----------|
| AED1                  | 8.231678 |
| PI                    | 7.348425 |
| IA                    | 7.928744 |
| MD                    | 6.584507 |
| AL                    | 7.492730 |
| Disciplinas Trancadas | 3.000000 |
| Total de Os           | 0.000000 |
| Faltas                | 0.000000 |
| CR                    | 2.261387 |

Name: 0, dtype: float64

Aprovado

# Decision Tree

- Gini index
- Profundidade máxima: 4
- Qtd mínima de amostras em nó folha: 6
  - Cross validation: 10
  - Score médio: 0,885

# Decision Tree

## Gini index

Dataset D, Atributo A que pode ser True ou False.

Valor 1 (D, A = True) =  $1 - (\text{acertos} / \text{total true})^2 - (\text{erros} / \text{total true})^2 = \text{valor 1}$

Valor 2 (D, A = False) =  $1 - (\text{acertos} / \text{total false})^2 - (\text{erros} / \text{total false})^2 = \text{valor 2}$

Gini index =  $(\text{total true} / \text{total}) \times \text{valor 1} + (\text{total false} / \text{total}) \times \text{valor 2}$

Melhor caso: 0

Pior caso: 0.5



