



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
DEPARTAMENTO DE COMPUTAÇÃO E AUTOMAÇÃO

DCA0133 - Aprendizagem de Máquina e Mineração de Dados

Aluno: Fernando Lucas Sousa Silva

Atividade avaliativa – Unidade 1 – Questão 2

2. Um determinado banco deve decidir se um cliente deve ou não receber um empréstimo bancário em função da sua condição de bom ou mau pagador. Considerando os dados de treinamento abaixo, aplique o classificador, no caso uma árvore de decisão, para atribuir a classe (rótulo) para o registro 12:

Registro	Tem casa própria	Estado Civil	Rendimentos	Bom Pagador
1	Sim	Solteiro	Alto	Não
2	Não	Casado	Médio	Não
3	Não	Solteiro	Baixo	Não
4	Sim	Casado	Alto	Não
5	Não	Divorciado	Médio	Sim
6	Não	Casado	Baixo	Não
7	Sim	Divorciado	Alto	Sim
8	Não	Solteiro	Médio	Sim
9	Não	Casado	Baixo	Não
10	Não	Solteiro	Médio	Sim
11	Sim	Divorciado	Médio	Não
12	Não	Divorciado	Alto	?

R: O método de Árvore de decisão é baseado em calcular o índice de Gini para encontrar o melhor valor que divide os nós da árvore. Desse modo, deve ser calculado para cada evidência do conjunto de dados. Assim, três etapas são realizadas: o primeiro passo é calcular o índice de Gini de cada evidência; o segundo passo é construir a árvore de decisão; o terceiro passo é realizar os dois passos anteriores novamente até que se tenha o Gini = 0.

A fórmula do Índice de Gini é: $Gini(D) = 1 - \sum_{i=1}^{n_{classe}} p_{(i)}^2$

- Passo 1:

$$\begin{aligned} \circ Gini(Tem\ casa\ própria) &= \left(\frac{4}{11} * \left(1 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2\right)\right) + \left(\frac{7}{11} * \right. \\ &\quad \left. \left(1 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2\right)\right) = 0,448 \end{aligned}$$

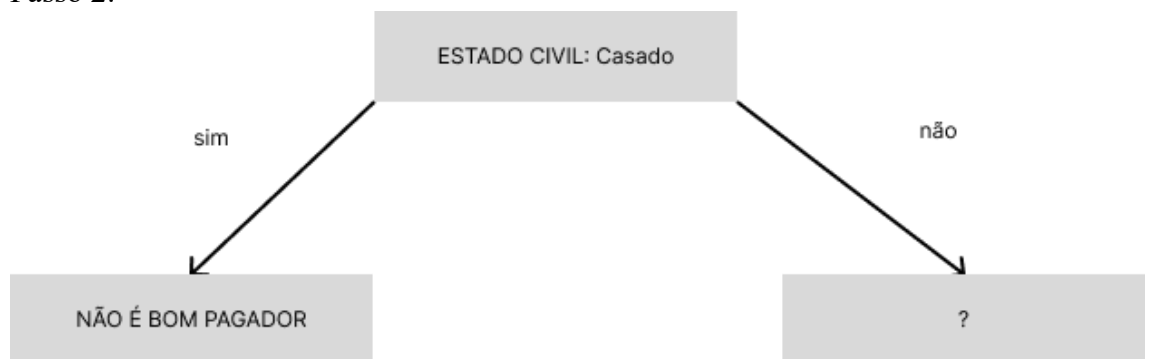
$$\begin{aligned} \circ Gini(Estado\ Civil: Solteiro) &= \left(\frac{4}{11} * \left(1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2\right)\right) + \left(\frac{7}{11} * \right. \\ &\quad \left. \left(1 - \left(\frac{2}{7}\right)^2 - \left(\frac{5}{7}\right)^2\right)\right) = 0,442 \end{aligned}$$



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
DEPARTAMENTO DE COMPUTAÇÃO E AUTOMAÇÃO

- $Gini (Estado Civil: Casado) = (\frac{4}{11} * (1 - (\frac{0}{4})^2 - (\frac{4}{4})^2)) + (\frac{7}{11} * (1 - (\frac{4}{7})^2 - (\frac{3}{7})^2)) = 0,312$
- $Gini (Estado Civil: Divorciado) = (\frac{3}{11} * (1 - (\frac{2}{3})^2 - (\frac{1}{3})^2)) + (\frac{8}{11} * (1 - (\frac{2}{8})^2 - (\frac{6}{8})^2)) = 0,394$
- $Gini (Rendimentos: Alto) = (\frac{3}{11} * (1 - (\frac{1}{3})^2 - (\frac{2}{3})^2)) + (\frac{8}{11} * (1 - (\frac{3}{8})^2 - (\frac{5}{8})^2)) = 0,462$
- $Gini (Rendimentos: Médio) = (\frac{5}{11} * (1 - (\frac{3}{5})^2 - (\frac{2}{5})^2)) + (\frac{6}{11} * (1 - (\frac{1}{6})^2 - (\frac{5}{6})^2)) = 0,420$
- $Gini (Rendimentos: Baixo) = (\frac{3}{11} * (1 - (\frac{0}{3})^2 - (\frac{3}{3})^2)) + (\frac{8}{11} * (1 - (\frac{4}{8})^2 - (\frac{4}{8})^2)) = 0,364$

- Passo 2:



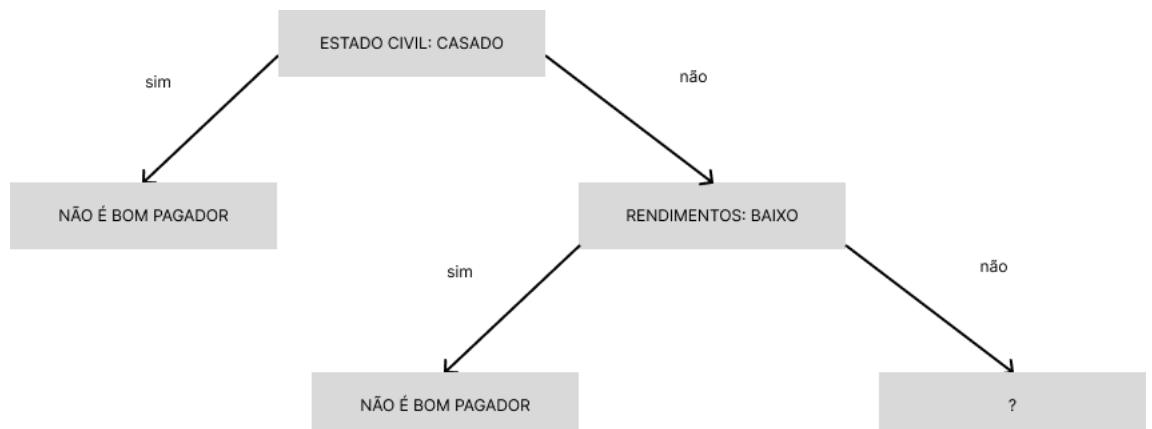
- 2ª Iteração - Passo 1 – Calcular o índice de Gini novamente desconsiderando as amostras de Estado Civil que respondem como “Casado”:



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
DEPARTAMENTO DE COMPUTAÇÃO E AUTOMAÇÃO

Registro	Tem casa própria	Estado Civil	Rendimentos	Bom Pagador
1	Sim	Solteiro	Alto	Não
2	Não	Casado	Médio	Não
3	Não	Solteiro	Baixo	Não
4	Sim	Casado	Alto	Não
5	Não	Divorciado	Médio	Sim
6	Não	Casado	Baixo	Não
7	Sim	Divorciado	Alto	Sim
8	Não	Solteiro	Médio	Sim
9	Não	Casado	Baixo	Não
10	Não	Solteiro	Médio	Sim
11	Sim	Divorciado	Médio	Não
12	Não	Divorciado	Alto	?

- $Gini(Tem\ casa\ própria) = (\frac{3}{7} * (1 - (\frac{1}{3})^2 - (\frac{2}{3})^2)) + (\frac{4}{7} * (1 - (\frac{3}{4})^2 - (\frac{1}{4})^2)) = 0,404$
- $Gini(Estado\ Civil: Solteiro | Divorciado) = (\frac{4}{7} * (1 - (\frac{2}{4})^2 - (\frac{2}{4})^2)) + (\frac{3}{7} * (1 - (\frac{2}{3})^2 - (\frac{1}{3})^2)) = 0,476$
- $Gini(Rendimentos: Alto) = (\frac{2}{7} * (1 - (\frac{1}{2})^2 - (\frac{1}{2})^2)) + (\frac{5}{7} * (1 - (\frac{3}{5})^2 - (\frac{2}{5})^2)) = 0,486$
- $Gini(Rendimentos: Médio) = (\frac{4}{7} * (1 - (\frac{3}{4})^2 - (\frac{1}{4})^2)) + (\frac{3}{7} * (1 - (\frac{1}{3})^2 - (\frac{2}{3})^2)) = 0,405$
- $Gini(Rendimentos: Baixo) = (\frac{1}{7} * (1 - (\frac{0}{1})^2 - (\frac{1}{1})^2)) + (\frac{6}{7} * (1 - (\frac{4}{6})^2 - (\frac{2}{6})^2)) = 0,381$
- 2ª Iteração – Passo 2 – Construir a árvore de decisão:





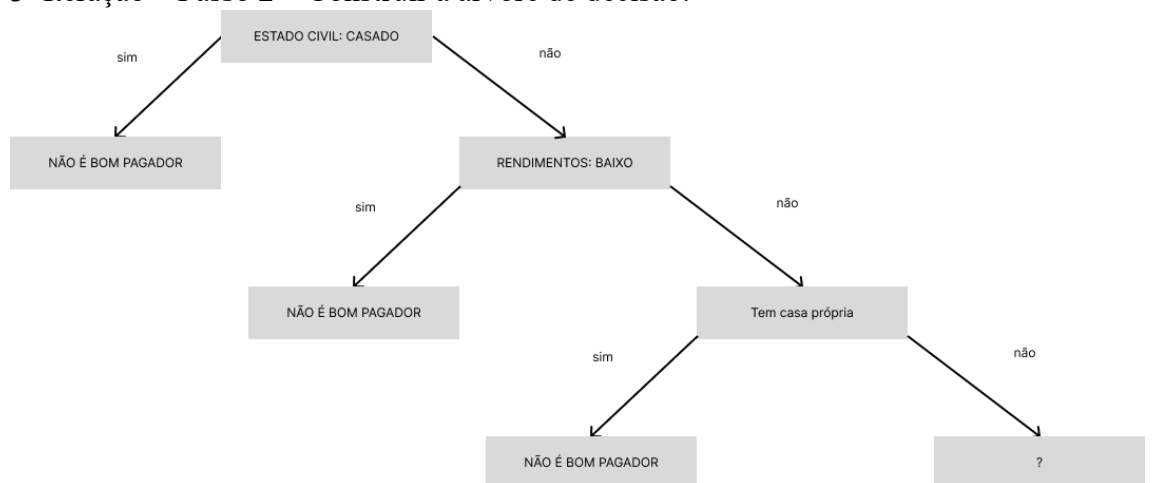
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
DEPARTAMENTO DE COMPUTAÇÃO E AUTOMAÇÃO

- 3ª Iteração – Passo 1 - Calcular o índice de Gini novamente desconsiderando as amostras de Rendimentos que respondem como “Baixo”:

Registro	Tem casa própria	Estado Civil	Rendimentos	Bom Pagador
1	Sim	Solteiro	Alto	Não
2	Não	Casado	Médio	Não
3	Não	Solteiro	Baixo	Não
4	Sim	Casado	Alto	Não
5	Não	Divorciado	Médio	Sim
6	Não	Casado	Baixo	Não
7	Sim	Divorciado	Alto	Sim
8	Não	Solteiro	Médio	Sim
9	Não	Casado	Baixo	Não
10	Não	Solteiro	Médio	Sim
11	Sim	Divorciado	Médio	Não
12	Não	Divorciado	Alto	?

- $Gini (Tem\ casa\ própria) = (\frac{3}{6} * (1 - (\frac{1}{3})^2 - (\frac{2}{3})^2)) + (\frac{3}{6} * (1 - (\frac{3}{3})^2 - (\frac{0}{3})^2)) = 0,222$
- $Gini (Estado\ Civil: Solteiro | Divorciado) = (\frac{3}{6} * (1 - (\frac{2}{3})^2 - (\frac{1}{3})^2)) + (\frac{3}{6} * (1 - (\frac{2}{3})^2 - (\frac{1}{3})^2)) = 0,444$
- $Gini (Rendimentos: Alto | Médio) = (\frac{2}{6} * (1 - (\frac{1}{2})^2 - (\frac{1}{2})^2)) + (\frac{4}{6} * (1 - (\frac{3}{4})^2 - (\frac{1}{4})^2)) = 0,417$

- 3ª Iteração – Passo 2 – Construir a árvore de decisão:





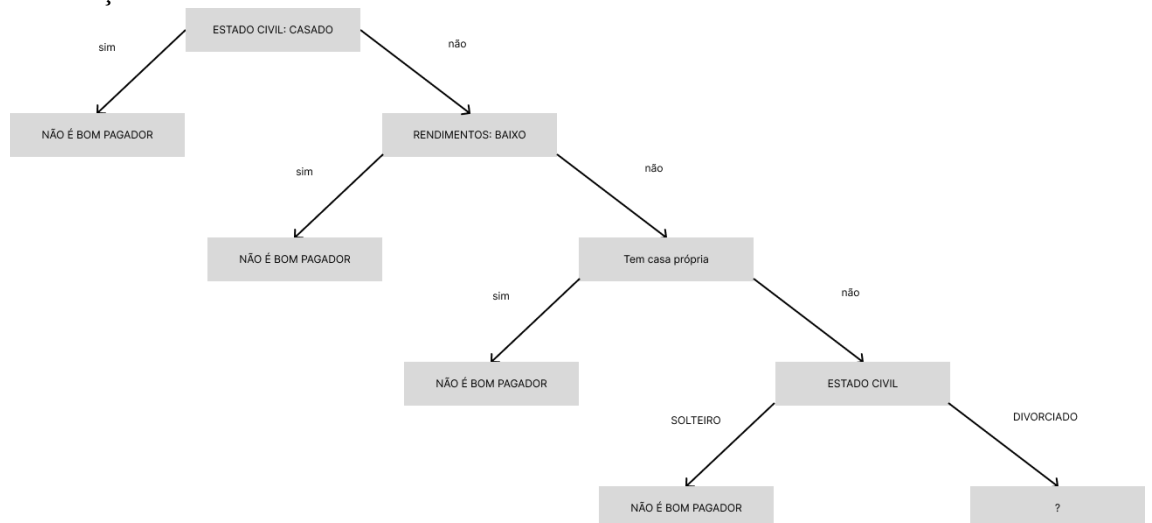
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
DEPARTAMENTO DE COMPUTAÇÃO E AUTOMAÇÃO

- 4ª Iteração – Passo 1 - Calcular o índice de Gini novamente desconsiderando as amostras de Tem casa própria que respondem como “Não”:

Registro	Tem casa própria	Estado Civil	Rendimentos	Bom Pagador
1	Sim	Solteiro	Alto	Não
2	Não	Casado	Médio	Não
3	Não	Solteiro	Baixo	Não
4	Sim	Casado	Alto	Não
5	Não	Divorciado	Médio	Sim
6	Não	Casado	Baixo	Não
7	Sim	Divorciado	Alto	Sim
8	Não	Solteiro	Médio	Sim
9	Não	Casado	Baixo	Não
10	Não	Solteiro	Médio	Sim
11	Sim	Divorciado	Médio	Não
12	Não	Divorciado	Alto	?

- $Gini (Tem\ casa\ própria) = (\frac{3}{3} * (1 - (\frac{1}{3})^2 - (\frac{2}{3})^2)) = 0,444$
- $Gini (Estado\ Civil: Solteiro | Divorciado) = (\frac{1}{3} * (1 - (\frac{0}{1})^2 - (\frac{1}{1})^2)) + (\frac{2}{3} * (1 - (\frac{1}{2})^2 - (\frac{1}{2})^2)) = 0,333$
- $Gini (Rendimentos: Alto | Médio) = (\frac{2}{3} * (1 - (\frac{1}{2})^2 - (\frac{1}{2})^2)) + (\frac{1}{3} * (1 - (\frac{0}{1})^2 - (\frac{1}{1})^2)) = 0,333$

- 4ª Iteração – Passo 2 – Construir a árvore de decisão:



- 5ª Iteração – Passo 1 - Calcular o índice de Gini novamente desconsiderando as amostras de Estado Civil que respondem como “Solteiro”:

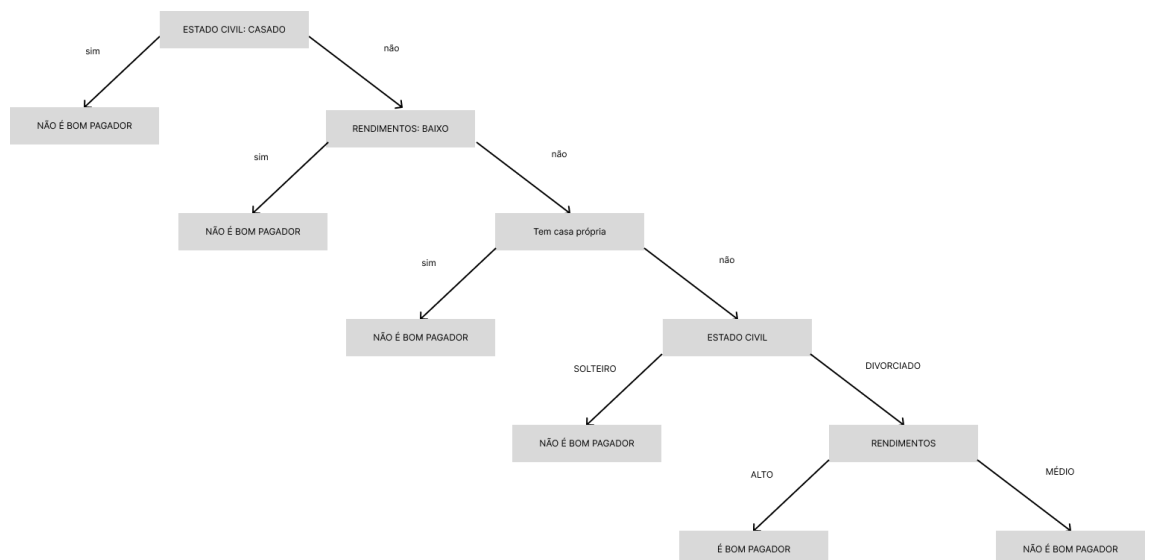


UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
DEPARTAMENTO DE COMPUTAÇÃO E AUTOMAÇÃO

Registro	Tem casa própria	Estado Civil	Rendimentos	Bom Pagador
1	Sim	Solteiro	Alto	Não
2	Não	Casado	Médio	Não
3	Não	Solteiro	Baixo	Não
4	Sim	Casado	Alto	Não
5	Não	Divorciado	Médio	Sim
6	Não	Casado	Baixo	Não
7	Sim	Divorciado	Alto	Sim
8	Não	Solteiro	Médio	Sim
9	Não	Casado	Baixo	Não
10	Não	Solteiro	Médio	Sim
11	Sim	Divorciado	Médio	Não
12	Não	Divorciado	Alto	?

- $Gini(Tem\ casa\ própria) = (\frac{2}{2} * (1 - (\frac{1}{2})^2 - (\frac{1}{2})^2)) = 0,5$
- $Gini(Rendimentos: Alto | Médio) = (\frac{1}{2} * (1 - (\frac{1}{1})^2 - (\frac{0}{1})^2)) + (\frac{1}{2} * (1 - (\frac{0}{1})^2 - (\frac{1}{1})^2)) = 0$

- 5ª Iteração – Passo 2 – Construir a árvore de decisão:





UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
DEPARTAMENTO DE COMPUTAÇÃO E AUTOMAÇÃO

- Conclusão
 - Como constatado durante as 5 iterações e a construção da árvore de decisão final, podemos dizer que por esse método e considerando esse conjunto de dados apresentados, temos que um cliente que não possui casa própria, é divorciado e possui rendimentos altos, é um bom pagador.