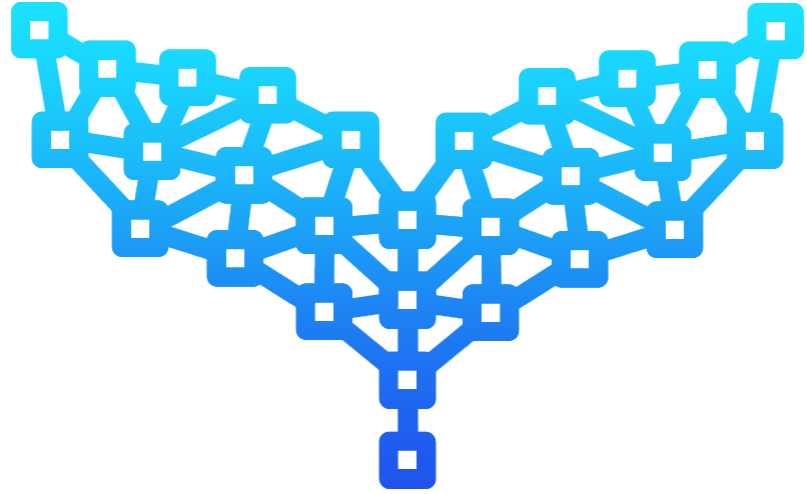
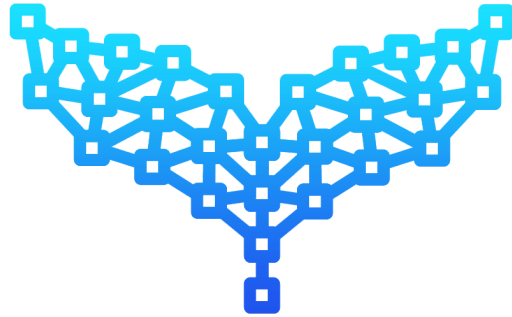


TAIL

Technology and Artificial
Intelligence League



DECISION TREE





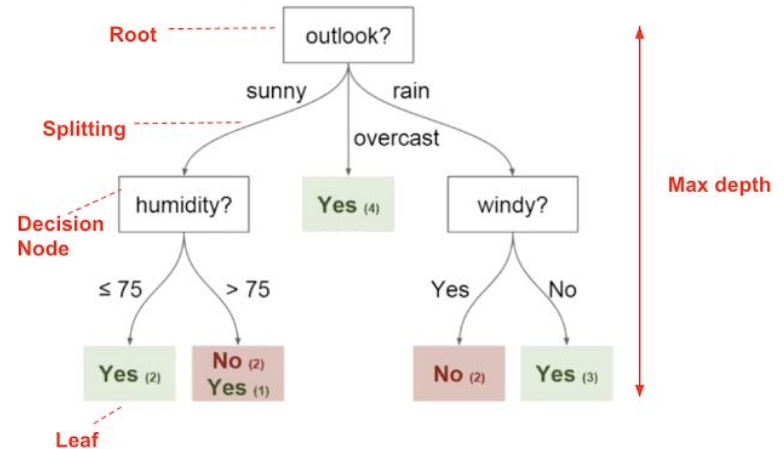
1. Introdução

A Árvore de Decisão é um modelo de Machine Learning usado para classificação ou para regressão. O objetivo desse método é dividir um problema em subseções consecutivas (fluxo de processamento recursivo).

Alguns exemplos em que pode ser aplicado:

- Resolvendo problemas que focam em prever chance de algum evento ocorrer (ex: desenvolvimento de diabetes);
- São muito utilizadas para análise de decisão.

Decision Tree Diagram

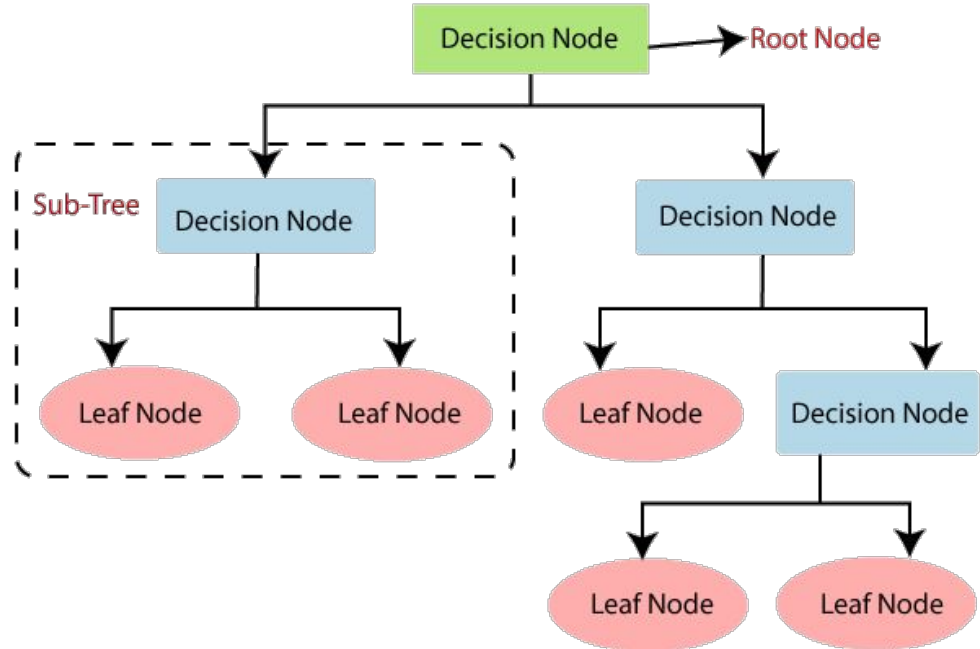


Exemplo gráfico do Decision Tree.

2. Componentes



- Nó interno (decision node);
- Folha (leaf node);
- Nó raiz (root node);
- Divisão (splitting);
- Ramificação (branch);
- Poda (prunning).





3. Entropia

O algoritmo faz uso da fórmula de Entropia e Ganho de Informação para definir os níveis de cada atributo previsor.

$$Entropy = \sum_{i=1}^C -p_i * \log_2(p_i)$$

Fórmula da Entropia

$$Gain(S, D) = H(S) - \sum_{V \in D} \frac{|V|}{|S|} H(V)$$

Fórmula de Ganho de Informação

4. Exemplo



Considere a seguinte tabela como exemplo:

Salário	Localização	Função	Decisão
alto	longe	interessante	SIM
baixo	perto	desinteressante	NÃO
baixo	longe	interessante	SIM
alto	longe	desinteressante	NÃO
alto	perto	interessante	SIM
baixo	longe	desinteressante	NÃO

Tabela de predição de aceite ou não do emprego

4. Exemplo



Considere os seguintes cálculos de Entropia e Ganho de Informação:

$$\text{Ganho(Decisão, Salário)} = 1 - (3/6 * 0,918 + 3/6 * 0,918) = 0,082$$

$$\text{Ganho(Decisão, Localização)} = 1 - (4/6 * 1,0 + 2/6 * 1,0) = 0,0$$

$$\text{Ganho(Decisão, Função)} = 1 - (3/6 * 0,0 + 3/6 * 0,0) = 1,0$$

$$\text{Entropia(alto)} = -(2/3) * \log_2(2/3) - (1/3) * \log_2(1/3) = 0,918$$

$$\text{Entropia(baixo)} = -(1/3) * \log_2(1/3) - (2/3) * \log_2(2/3) = 0,918$$

$$\text{Entropia(longe)} = -(2/4) * \log_2(2/4) - (2/4) * \log_2(2/4) = 1,0$$

$$\text{Entropia(perto)} = -(1/2) * \log_2(1/2) - (1/2) * \log_2(1/2) = 1,0$$

$$\text{Entropia(desinteressante)} = -(3/3) * \log_2(3/3) = 0,0$$

$$\text{Entropia(interessante)} = -(3/3) * \log_2(3/3) = 0,0$$

$$\text{Entropia(decisão)} = -(1/2) * \log_2(1/2) - (1/2) * \log_2(1/2) = 1,0$$

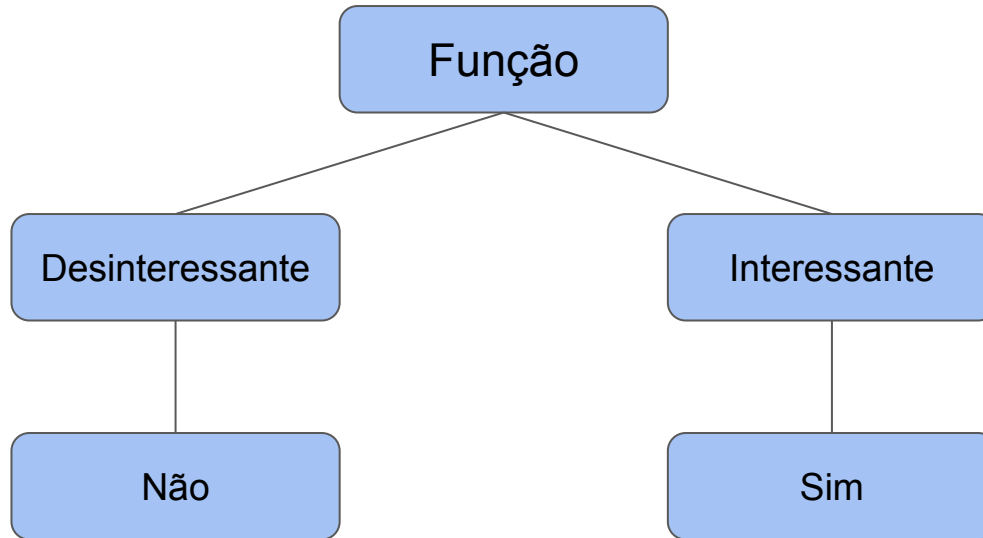
O atributo com maior Ganho de Informação fica no topo da árvore.

A entropia, por sua vez, mede o grau de impureza dos valores dos atributos.

4. Exemplo



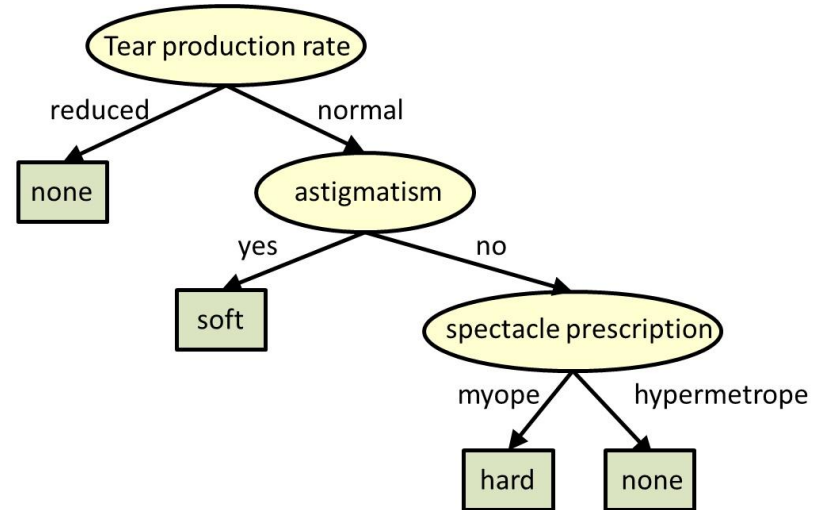
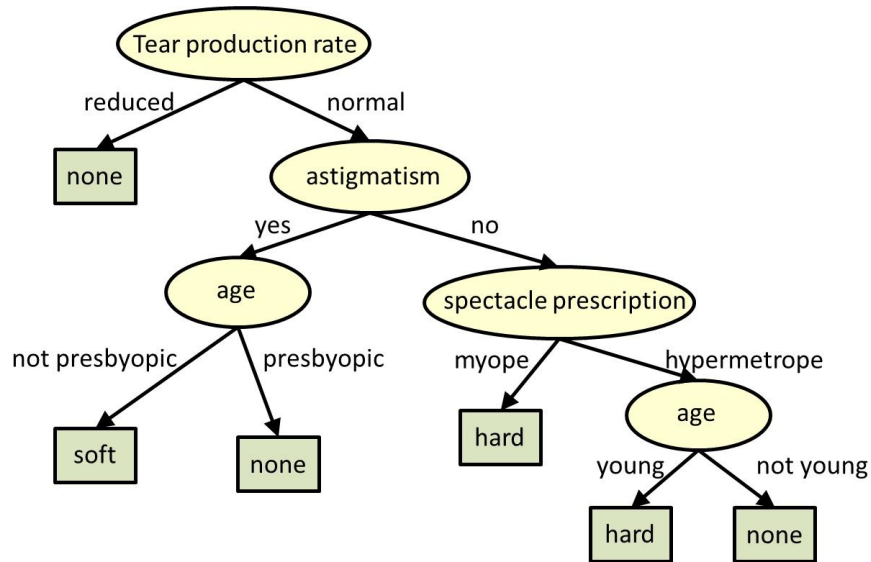
A partir dos cálculos podemos criar a seguinte árvore:



5. Podagem



A poda é um método de exclusão de subconjuntos da árvore, com o intuito de minimizar o erro do modelo.



Referências



LAURETTO, Marcelo S.. Árvores de Decisão. Disponível em:
<https://edisciplinas.usp.br/pluginfile.php/4469825/mod_resource/content/1/ArvoresDecisao_normalsize.pdf>. Acesso em 05 de Outubro de 2020;

ZUBEN, Von; ATTUX, Romis R. F.. Árvores de Decisão. Disponível em:
<ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia004_1s10/notas_de_aula/topico7_IA004_1s10.pdf>.
Acesso em 05 de Outubro de 2020;

CHOWDARY, Davuluri Hemanth. Decision Trees Explained With a Practical Example. Disponível em:
<<https://towardsai.net/p/programming/decision-trees-explained-with-a-practical-example-fe47872d3b53>>.
Acesso em 05 de Outubro de 2020.