

DECISION TREE LEARNING



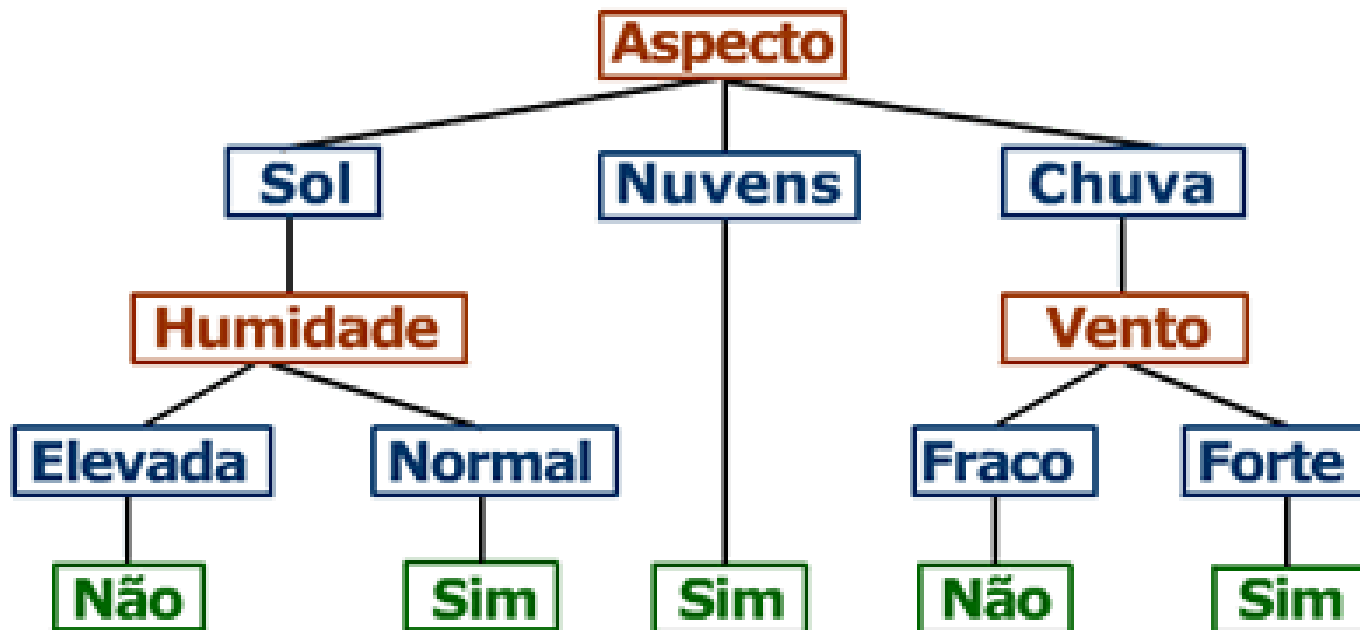
Leonardo e Maurício

O que é uma árvore de decisão?

- ⦿ Nodo raiz(estágio inicial do problema)
- ⦿ Condições aplicadas
- ⦿ Resultado da decisão em nodos folhas

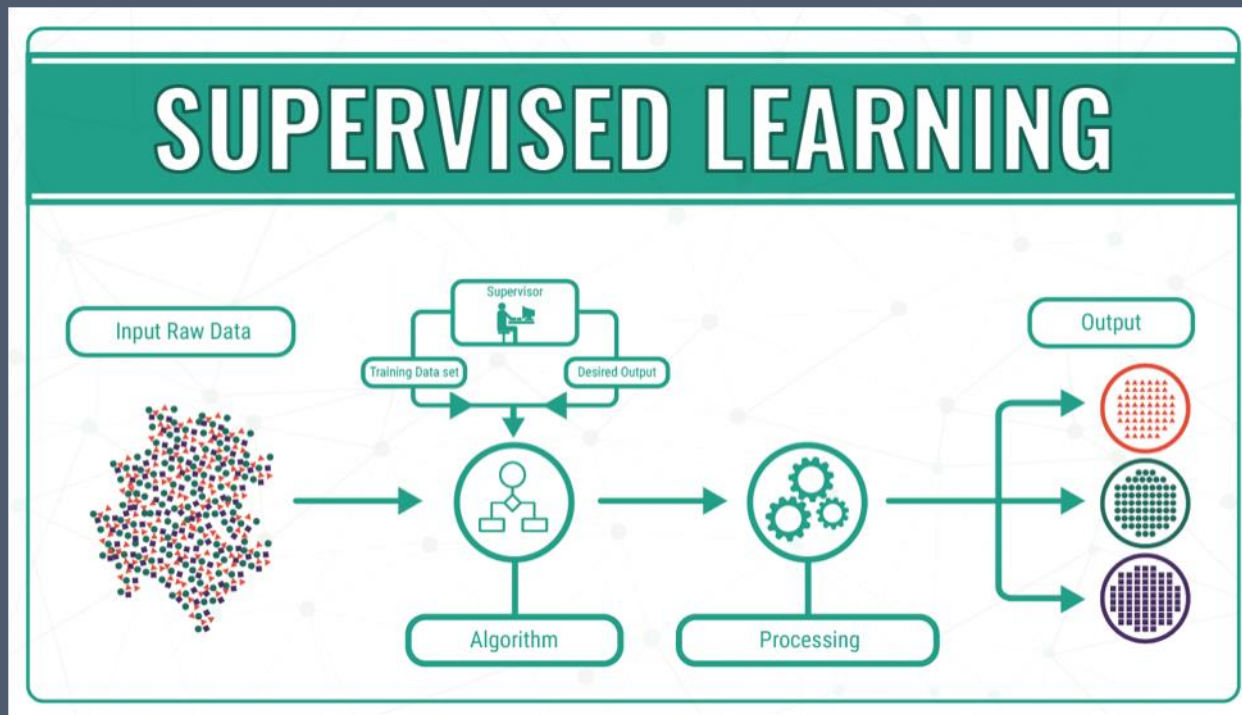
O que é uma árvore de decisão?

Árvore de Decisão para Jogar Tênis



Como funciona?

- Aprendizado supervisionado
- Exemplos de treinamento



Classificação de Decision Tree

- Classification Tree: Quando a predição do modelo nos mostra a qual classe o dado pertence
- Regression Tree: Quando a predição do modelo é um número ou valor

Análise da Técnica - Prós

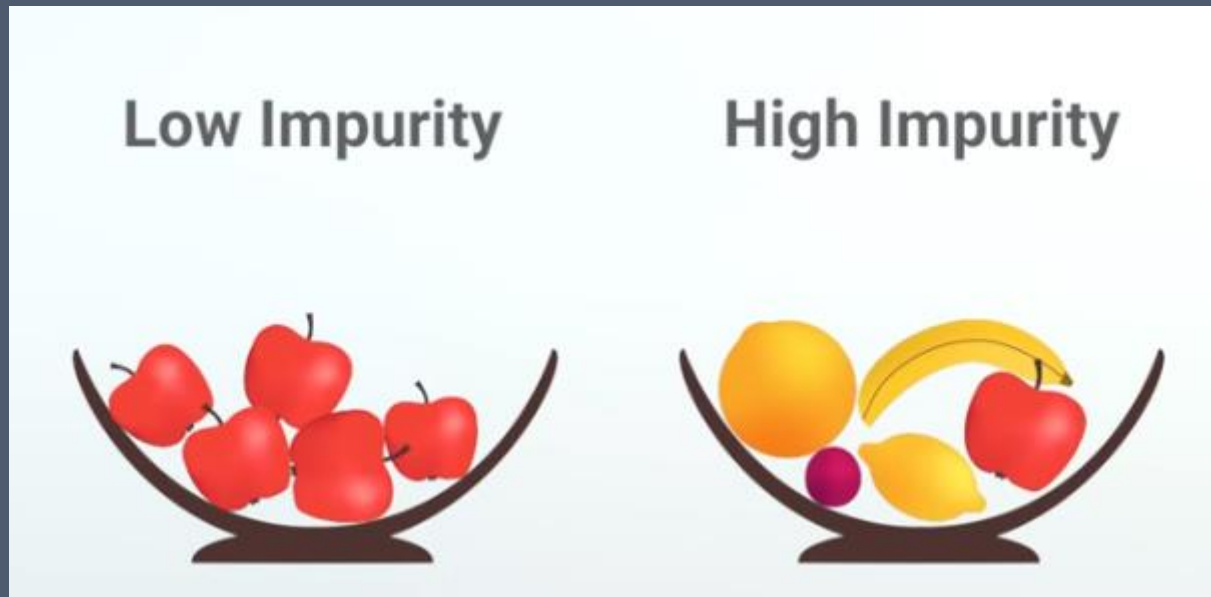
- De fácil implementação
- Pode ser usada com entradas tanto numéricas quanto categóricas

Análise da Técnica - Cons

- ⦿ Não funciona bem quando existem muitas variáveis não relacionadas
- ⦿ Pode ser instável, uma pequena mudança nos dados, e conseqüentemente nas previsões
- ⦿ (Overfitting) Pode criar árvore muito complexas, para resolver problemas simples

Métricas - Gini Impurity

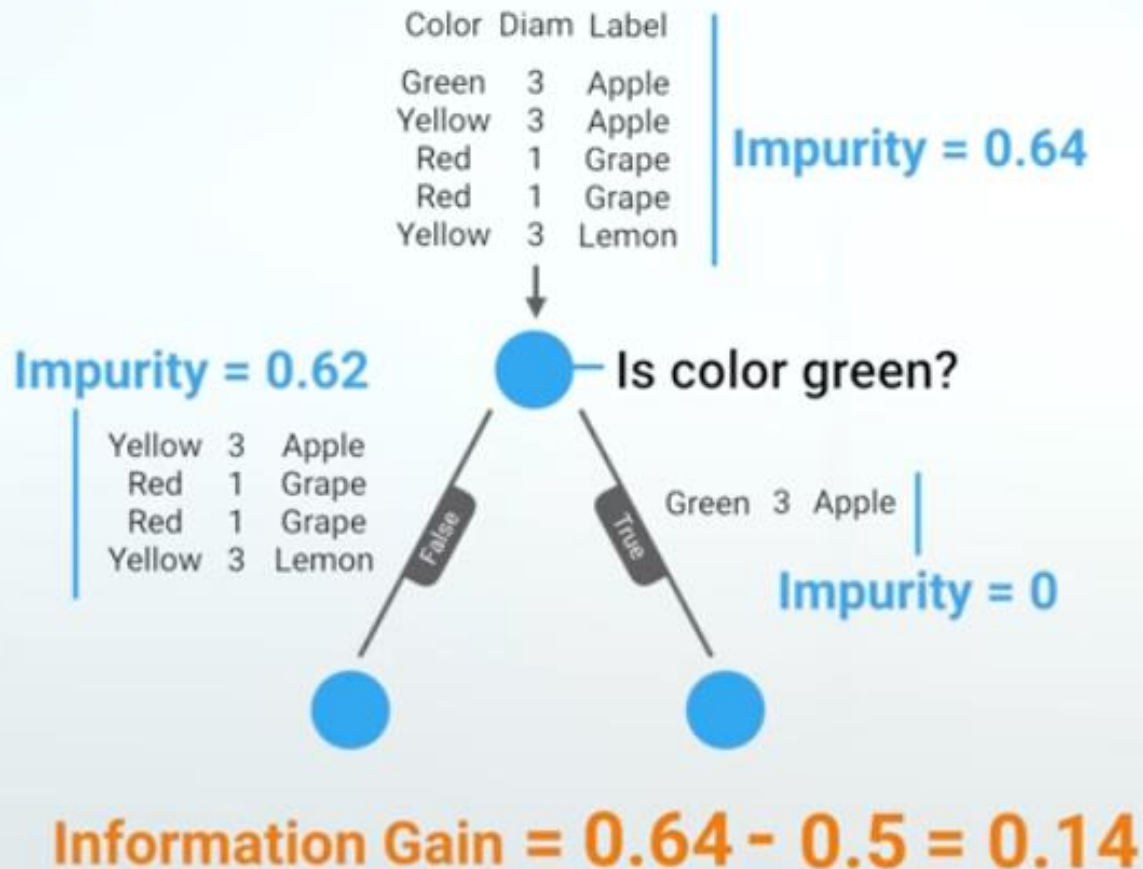
- Gini Impurity é a medida de com que frequência um elemento escolhido em um set vai ser incorretamente identificado pelo seu subset



Métricas – Ganho de Informação

- É a métrica que permite encontrar a questão que mais diminui a incerteza
- Uma métrica que ajuda a encontrar a pergunta que tem o melhor potencial para desembaralhar e agrupar os sets

Formação da árvore



$$\text{Avg Impurity} = \frac{4}{5} * 0.62 + \frac{1}{5} * 0 \approx 0.5$$

Color	Diam	Label
Green	3	Apple
Yellow	3	Apple
Red	1	Grape
Red	1	Grape
Yellow	3	Lemon



Information Gain

Question	Gain
Color == Green?	0.14
Diameter >= 3?	0.37
Color == Yellow?	0.17
Color == Red?	0.37
Diameter >= 1?	0

Técnicas de melhoria - Pruning

- Algoritmo para melhorias de decision tree que permite a eliminação de nodos que não contribuem para a classificação do modelo

Algoritmos Notáveis de Decision Tree Learning

- ◉ ID3 (Iterative Dichotomiser 3)
- ◉ C4.5 (sucessor do ID3) Melhorias em relação ao ID3:
 1. Lida com valores discretos e contínuos;
 2. Lida com data com valores faltando;
 3. Já implementa Pruning Tree;
- ◉ CART (Classification And Regression Tree)
- ◉ CHAID (CHi-squared Automatic Interaction Detector). Executa divisões multi-nível quando processa árvores de classificação
- ◉ MARS: Melhora o funcionamento de árvores de decisões com dados numéricos

Referências

- <https://heartbeat.fritz.ai/introduction-to-decision-tree-learning-cd604f85e236>
- <https://www.oreilly.com/library/view/machine-learning-with/9781787121515/697c4c5f-1109-4058-8938-d01482389ce3.xhtml>
- https://en.wikipedia.org/wiki/Decision_tree_learning#Advantages
- https://en.wikipedia.org/wiki/Decision_tree_learning
- <https://www.youtube.com/watch?v=LDRbO9a6XPU>
- https://en.wikipedia.org/wiki/ID3_algorithm
- https://en.wikipedia.org/wiki/C4.5_algorithm