

Decision Tree



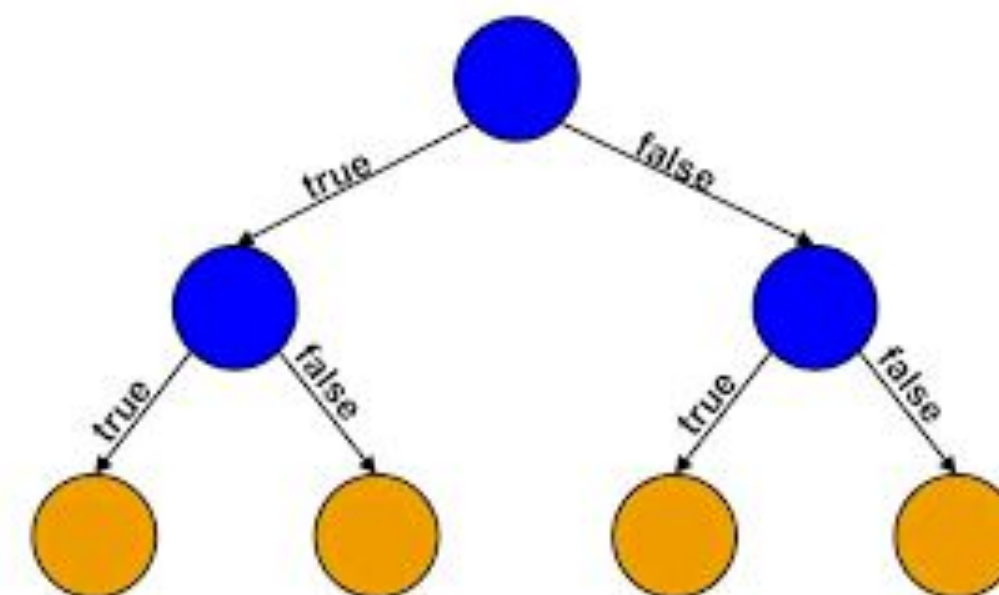
Por que?

Algoritmo que consegue fazer a classificação de decisões a medida que temos dados que podem ser classificados a partir da sua entrada.

Matematicamente como acontece?

Calculamos a ordem que as decisões aparecem na árvore através do índice de Gini. E assim quanto menor o índice, mais acima esse nó deve aparecer na árvore de decisão.

$$G = \sum_{i=1}^C p(i) * (1 - p(i))$$





Vantagens e desvantagens

As vantagens deste algoritmo é que é muito simples de implementar, entender e visualizar a aplicação. Além de não precisar de todos os dados normalizados.

E uma das desvantagens é porque caso tenhamos muitos dados no nosso dataset, ele pode demorar no treinamento, visto que precisa calcular o índice milhares de vezes para decidir a ordem dos nós.

Na prática

Na prática, a biblioteca Scikit Learn possui modelos prontos para serem treinados, possuindo uma documentação bem escrita para estudo.