



**escola
britânica de
artes criativas
& tecnologia**

**Profissão Cientista de Dados
Árvore de decisão**

Prós e contras

Prós x contras

Algumas vantagens das árvores de decisão são:

- Simples de entender e interpretar. As árvores podem ser visualizadas.
- Requer pouca preparação de dados. Outras técnicas geralmente requerem normalização de dados, variáveis dummy (flag) precisam ser criadas e valores em branco precisam ser removidos.
- Capaz de lidar com dados numéricos e categóricos. Outras técnicas geralmente são especializadas na análise de conjuntos de dados que possuem apenas um tipo de variável.
- Capaz de lidar com problemas de múltiplas saídas.
- Usa um modelo de caixa branca. Se uma determinada situação é observável em um modelo, a explicação para a condição é facilmente explicada pela lógica booleana. Por outro lado, em um modelo de caixa preta (por exemplo, em uma rede neural), os resultados podem ser mais difíceis de interpretar.

As desvantagens das árvores de decisão incluem:

- As árvores de decisão podem ser tão complexas que não generalizam bem os dados. Isso é chamado de overfitting. Mecanismos como poda, definir o número mínimo de amostras necessárias numa folha ou definir a profundidade máxima da árvore são necessários para evitar este problema.
- As árvores de decisão podem ser instáveis, pois pequenas variações nos dados podem resultar na geração de uma árvore completamente diferente.
- As previsões das árvores de decisão não são suaves nem contínuas. Portanto, eles não são bons em extração.
- As árvores de decisão podem ser enviesadas se algumas classes dominam no conjunto de dados. Portanto, é recomendável equilibrar o conjunto de dados antes de ajustá-lo à árvore de decisão.