

Árvore de Decisão

Uma **Árvore de Decisão** é um modelo de aprendizagem supervisionada que é usado para classificação e regressão. Para além disso, uma Árvore de Decisão é um grafo hierarquizado em que: - O nodo raiz representa todo o dataset; - Cada ramo representa a seleção entre um conjunto de alternativas; - Cada folha representa uma decisão; - Cada nodo interno testa um atributo do dataset; - Cada ramo identifica um valor (ou conjunto de valores) do nodo testado;

Este modelo é considerado uma técnica de aprendizagem supervisionada, uma vez que é necessário um conjunto de dados de treino que contenha a resposta correta para cada exemplo.

Modelos de Decisão

Top-Down: - O modelo é construído a partir do conhecimento de especialistas; - O “todo” é dividido em “partes” menores;

Bottom-Up: - O modelo é construído a partir da identificação de relações no dataset; - O modelo é induzido por “generalização” dos dados;

As árvores de decisão seguem o paradigma *Bottom-Up*: - Toda a informação sobre cada item de dados deve estar definido numa coleção fixa e finita de atributos; - Deste modo, objetos distintos não podem requerer coleções distintas de atributos; - Quando o conjunto dos níveis de decisão é conhecido à priori, a construção do modelo segue um paradigma de aprendizagem supervisionado; - Quando o conjunto dos níveis de decisão é calculado pelo modelo, a sua construção segue um paradigma de aprendizagem não supervisionado;

Para além disso, as árvores de decisão podem ser: - **Contínuas:** Quando os atributos representam um conjunto ou intervalos de possíveis valores e as folhas da árvore representam valores contínuos; - **Discretas:** Quando os atributos representam uma categoria ou classe de valores e as folhas da árvore representam valores discretos;

Vantagens e Desvantagens

Vantagens: - Simples de entender e interpretar; - Requer pouca preparação de dados; - Pode lidar com dados numéricos e categóricos;

Desvantagens: - Tendência para *overfitting*: Pode criar árvores muito complexas que não generalizam bem os dados; - Instabilidade: Pequenas variações nos dados podem resultar em árvores muito diferentes; - Enviesamento: Árvores de decisão podem ser enviesadas para atributos com mais valores;

Referências

- Decision Tree: Important things to know