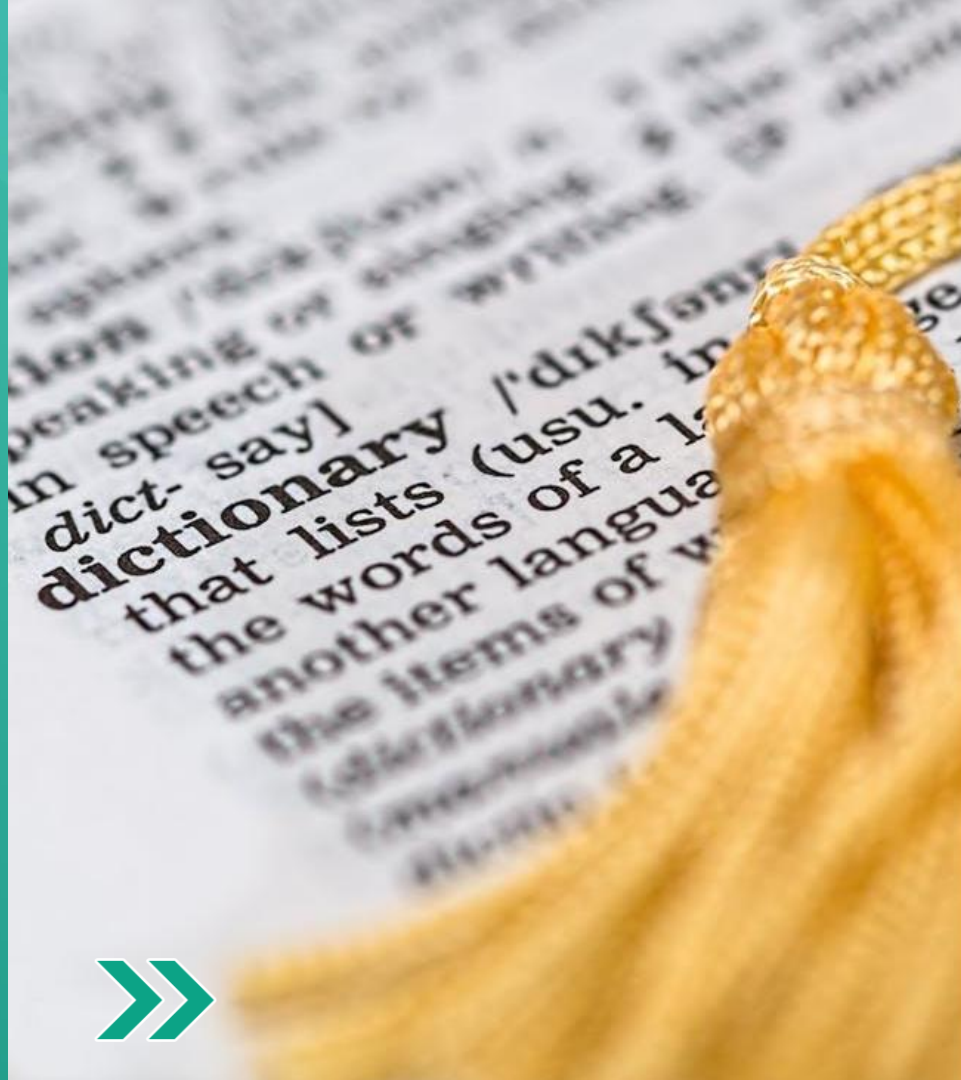
escola britânica de artes criativas & tecnologia

Profissão: Cientista de Dados

escola britânica de artes criativas & tecnologia

GLOSSÁRIO



Árvores I

- Entenda a árvore de decisão
- Carregue e trate dados
- Saiba o melhor valor e variável para quebra
- Domine a poda da árvore
- Acompanhe os pós e contras

Dica: para encontrar rapidamente a palavra que procura aperte o comando CTRL+F e digite o termo que deseja achar.

Entenda a árvore de decisão

Entenda a árvore de decisão

- **Algoritmos supervisionados**
  Categoria de algoritmos de inteligência artificial que inclui árvores de decisão. São divididos em dois grupos: classificação e regressão.
- **Classificação**
  Usada quando o resultado é binário, como bom ou mau. Um exemplo é determinar se um e-mail é spam ou não.
- **Regressão**
  Usada quando o resultado é contínuo, como prever o preço de um item.
- **Nó raiz**
  O ponto de partida de uma árvore de decisão. Tem uma profundidade de zero.

Entenda a árvore de decisão

Ramos

As "linhas" que conectam os nós em uma árvore de decisão.

Nós Internos

Nós que têm pelo menos um nó filho em uma árvore de decisão.

Folhas

Nós que não têm filhos em uma árvore de decisão. Eles representam a decisão final tomada após percorrer a árvore.

Profundidade

Refere-se ao tamanho de uma árvore de decisão. A raiz tem uma profundidade de zero e cada camada subsequente aumenta em um.

Carregue e trate dados

Carregue e trate dados

- **drop_duplicates**
  É um método em Python usado para remover linhas duplicadas de um DataFrame.

- **Reindexação**
  É o processo de alterar o índice de um DataFrame para que ele corresponda ao número de linhas.

- **isnull**
  É um método em Python usado para identificar valores ausentes em um DataFrame.

- **get_dummies**
  Método em Python usado para converter variáveis categóricas em numéricas, criando uma nova coluna para cada categoria e usando 0s e 1s para indicar a presença de cada categoria.

Carregue e trate dados

- **Variáveis explicativas (X) e variável alvo (y)**
  m aprendizado de máquina, as variáveis explicativas são as variáveis que são usadas para prever a variável alvo. A variável alvo é a variável que estamos tentando prever.

Saiba o melhor valor e variável para quebra

Saiba o melhor valor e variável para quebra

Classificador (clf)

É um termo usado em aprendizado de máquina para se referir a um algoritmo que é capaz de classificar dados não vistos ou futuros em uma classe específica com base em um modelo treinado em dados previamente vistos.

Impureza de Gini

É uma métrica usada para medir a qualidade de uma divisão em uma árvore de decisão. Quanto menor o valor de Gini, mais 'puro' é o nó, ou seja, os dados dentro do nó são principalmente de uma classe.

Saiba o melhor valor e variável para quebra

- **Scikit-learn**
  É uma biblioteca de aprendizado de máquina em Python que fornece uma seleção de algoritmos de aprendizado de máquina eficientes para uso em classificação, regressão e agrupamento, bem como ferramentas para pré-processamento de dados, seleção e avaliação de modelos.

Domine a poda da árvore

Domine a poda da árvore

Alfa (C alfa)

É um parâmetro usado na poda de árvores de decisão. Ele limita o número de nós na árvore. Quanto maior o C alfa, menos nós a árvore terá.

Ponto de corte

É o valor que separa os valores de uma variável contínua em dois grupos. A determinação do ponto de corte é um passo importante na construção de árvores de decisão.

Método 'complex'

É um método usado em árvores de decisão para retornar todos os C alfas.

Pós-poda

Método de poda de árvores de decisão que envolve o crescimento da árvore até um certo ponto e, em seguida, a identificação do melhor ponto de corte.

Acompanhe os pós e contras

Acompanhe os pós e contras

- **Caixa Branca**
  Em aprendizado de máquina, é um modelo cujo funcionamento interno é transparente e pode ser facilmente compreendido. As árvores de decisão são consideradas modelos de caixa branca.
- **Random Forest**
  É um algoritmo de aprendizado de máquina que usa um conjunto de árvores de decisão para fazer previsões. É baseado no conceito de árvores de decisão.
- **Problemas de Múltiplas Saídas**
  São problemas onde um modelo precisa prever mais de uma variável de saída. As árvores de decisão podem lidar com esses tipos de problemas.
- **XGBoost**
  É uma implementação otimizada do algoritmo de aumento de gradiente que é conhecido por sua velocidade e desempenho. É baseado no conceito de árvores de decisão.

escola britânica de artes criativas & tecnologia

Bons estudos!