

1. Qual a diferença entre Big Data e Ciência de Dados?

- Big Data lida com a tecnologia, gerenciando e processando os dados. Já a Ciência de Dados lida com a criação de soluções para modelagem de dados.

2. De que forma Estatística, Mineração de Dados e Aprendizagem de Máquina interagem com Ciência de Dados?

- São 3 áreas que se completam uma com a outra. Você precisa entender Estatística para poder calcular os dados Minerados e assim fazer deles um bom uso com Aprendizagem de Máquina.

3. Mostre a importância do conhecimento de domínio para o cientista de dados.

- Tendo um domínio maior sobre os dados estudados, tem-se a vantagem para entendê-los e manuseá-los com melhor precisão no objetivo esperado.

6. Cite 2 técnicas para remoção de ruídos e, para cada uma, mostre uma vantagem e uma desvantagem.

- Regressão:
 - Podendo ser linear ou múltipla.
- Binning:
 - Minimiza o efeito de pequenos erros de observação.

7. Qual é a importância de utilizar as seguintes abordagens de redução de dados no contexto de Ciência dos Dados?

- A. Redução de dimensionalidade
 - a. Reduzir o número de variáveis.
- B. Redução de numerosidade
 - a. Reduz o número de tuplas.

8. De que forma pode-se detectar overfitting em um classificador?

- Quando o modelo tem um ótimo desempenho no treino, mas ao utilizar os dados de teste o desempenho é péssimo.

9. Em quais tipos de problemas é preferível utilizar leave-one-out a utilizar K-fold cross-validation?

- Em pequenos conjuntos de dados.