

Oficina 8: Usando técnicas e estratégias de análise de dados

Aluno: Herik Douglas Oliveira Reinaldo

Vamos imaginar uma situação: você está analisando um conjunto de dados de clientes de um supermercado. Esse conjunto inclui várias variáveis, como idade, frequência de compras, categorias de produtos adquiridos e o total gasto. Agora, compartilhe suas ideias! Como você acredita que podemos simplificar essa análise para identificar padrões de compra? A empresa quer entender melhor o comportamento dos clientes, e esse é um desafio interessante! Quais estratégias ou técnicas você acha que poderíamos usar para tornar essa análise mais clara e eficaz? Estou curioso para suas ideias!

Pergunta:

1. Qual técnica de aprendizado não supervisionado você recomendaria para reduzir a dimensionalidade do conjunto de dados e qual é o objetivo principal dessa técnica?
2. Após a redução da dimensionalidade, a empresa deseja encontrar quais produtos são frequentemente comprados juntos. Qual método de associação seria mais adequado para essa tarefa e por quê?

Para ajudá-lo a se guiar na sua resposta, veja alguns aspectos que você pode levar em consideração:

1. Técnica de Redução de Dimensionalidade:
 - Nome da Técnica
 - Objetivo Principal da Técnica
 - Justificativa da escolha da técnica
2. Método de Associação:
 - Nome do Método
 - Justificativa para a escolha do método
 - Explicação de como o método atende ao objetivo de encontrar padrões de associação

1 - Qual técnica de aprendizado não supervisionado você recomendaria para reduzir a dimensionalidade do conjunto de dados e qual é o objetivo principal dessa técnica?

Nome da técnica: Análise de Componentes Principais (PCA)

Objetivo Principal da Técnica: Capturar a maior variabilidade dos dados com o menor número de componentes possíveis.

Justificativa da Escolha da Técnica: A aplicação do PCA nesse conjunto de dados pode reduzir as quatro variáveis originais (como idade, frequência de compras, categorias de produtos e total gasto) para dois componentes principais, que serão

combinações lineares das variáveis originais, ou seja, combinam essas informações de forma a preservar a maior parte da variabilidade dos dados. Esses componentes não são correlacionados entre si, o que facilita a visualização e análise dos padrões de comportamento dos clientes.

2 - Após a redução da dimensionalidade, a empresa deseja encontrar quais produtos são frequentemente comprados juntos. Qual método de associação seria mais adequado para essa tarefa e por quê?

Nome da técnica: Regras de associação com Apriori

Objetivo Principal da técnica: Encontrar itens frequentes e gerar regras de associação.

Explicação de como o método atende ao objetivo de encontrar padrões de associação: O método analisa o conjunto de dados e busca regras como “se o cliente compra x, também compra y”. Com essas regras, o supermercado pode fazer promoções inteligentes, organizar melhor os produtos nas prateleiras ou sugerir produtos no app.