

Teste Cognitivo.AI

Olá ! Obrigada pelo interesse em fazer parte da rede de experts da Cognitivo.ai! Esse é um teste com o objetivo de conhecer um pouco mais sobre a sua forma de trabalhar com problemas que envolvem engenharia de dados.

O problema

O presente problema se refere aos dados de vinhos portugueses "Vinho Verde", que possuem variantes de vinho branco e tinto. Devido a questões de privacidade, apenas variáveis físico-químicas (input) e sensoriais (output) estão disponíveis (por exemplo, não há dados sobre tipo de uva, marca do vinho, preço de venda, etc).

Objetivo

Criar um modelo para estimar a qualidade do vinho.

Informação sobre os atributos

Variáveis input (baseado em testes físico-químicos):

1. Tipo
2. Acidez fixa
3. Volatilidade da acidez
4. Ácido cítrico
5. Açúcar residual
6. Cloretos
7. Dióxido de enxofre livre
8. Dióxido de enxofre total
9. Densidade
10. pH
11. Sulfatos
12. Álcool

Variável output (baseado em dado sensorial):

13. Qualidade (score entre 0 and 10)

Resolução

1. Faça uma análise exploratória para avaliar a consistência dos dados e identificar possíveis variáveis que impactam na qualidade do vinho.
2. Para a realização deste teste você pode utilizar o software de sua preferência (Python ou R), só pedimos que compartilhe conosco o código fonte (caso queira utilizar um

repositório git, sinta-se à vontade). Além disso, inclua um arquivo README.md onde você deve cobrir as respostas para os 5 pontos abaixo:

- a. Como foi a definição da sua estratégia de modelagem?
- b. Como foi definida a função de custo utilizada?
- c. Qual foi o critério utilizado na seleção do modelo final?
- d. Qual foi o critério utilizado para validação do modelo? Por que escolheu utilizar este método?
- e. Quais evidências você possui de que seu modelo é suficientemente bom?

Descrição dos dados

https://drive.google.com/open?id=1-oG5-kBt9xQ3Li4PEexpIA9_7RZhRM1f