Disciplina: ***Machine Learning II***

Trabalho: ***Aplicação de Técnicas ML***

Um importador do Brasil deseja ter em seu canal de vendas (e-commerce) vinhos verdes de uma região de Portugal. Ele gostaria de entender um pouco mais sobre esses vinhos para ajudá-lo na compra(importação) mas também para dar suporte ao atendimento/ dúvidas de seus clientes.

As 3 etapas desse desafio serão baseadas na base de vinhos da região de Portugal com as variáveis de características físico-químicas (composição) dos vinhos.

<https://archive.ics.uci.edu/ml/datasets/wine+quality>

**Etapa 1:**

Quais são as características físico-químicas do vinho (branco) que impactam na nota de avalição sensorial (Quality) dos vinhos dessa região?

1. Definição da variável resposta
2. Análise Exploratória dos dados
3. Exclusão de *outliers*, caso necessário (sempre explicando a opção) e a quantidade final após a exclusão.
4. Amostra de Treinamento dos modelos e Validação
5. Padronização dos dados, caso necessário
6. Para cada etapa colocar comentários sobre a técnica utilizada, a saída do modelo e análise sobre as variáveis utilizadas e seus respectivos resultados
7. Com a métrica correta faça a comparação entre os diferentes algoritmos e qual foi o melhor entre eles.

**Etapa 2:**

Sabendo que os vinhos com notas>=6 são considerados vinhos de boa qualidade (para os vinhos brancos) como posso selecionar no momento da importação os vinhos de boa qualidade?

1. Definição da variável resposta
2. Análise Exploratória dos dados
3. Exclusão de *outliers*, caso necessário (sempre explicando a opção) e a quantidade final após a exclusão.
4. Amostra de Treinamento dos modelos e Validação
5. Padronização dos dados, caso necessário
6. Para cada etapa colocar comentários sobre a técnica utilizada, a saída do modelo e análise sobre as variáveis utilizadas e seus respectivos resultados
7. Com a métrica correta faça a comparação entre os diferentes algoritmos e qual foi o melhor entre eles.

**Etapa 3:**

Como paladar é uma questão “pessoal” o importador deseja trazer vinhos brancos dessa região, mas gostaria de ter uma variedade de tipos desses vinhos, como podemos auxiliá-los.

Alguns passos para auxiliar nessa etapa:

1. Critérios de seleção e definição das variáveis segmentadoras
2. Análise Exploratória dos dados
3. Avaliação de outliers, indicar a exclusão ou não
4. Critério de padronização e análise de correlação entre as variáveis
5. Critérios para definição da quantidade de grupos (clusters)
6. Resultados encontrados – Caracterização dos Clusters
7. Para cada etapa anterior colocar comentários sobre a técnica utiliza, as respectivas saídas e análise sobre as variáveis utilizadas e seus respectivos resultados.
8. Resultados encontrados – Centroides, Quantidade de vinhos por clusters e Caracterização dos Clusters
9. Cruzar os clusters obtidos a partir das variáveis segmentadoras e a nota dado ao vinho (Quality). Faça uma avaliação dos resultados por cluster.

**Etapa 4:**

Na análise realizada na Etapa3 faz sentido fazer uma redução da dimensionalidade dos dados?

1. Explicação da necessidade/adequação
2. Utilizando componentes principais (PCA), e selecione as variáveis adequadas para isso e após a aplicação do PCA explique sobre a quantidade de componentes que você escolheria e a interpretação dos componentes.

.