

1. Escolha uma base de dados. Essa base deve ser multivariada e possuir uma variável objetivo, que pode ser usada na fase de treino (modelos supervisionados) e/ou na validação do modelo (modelos supervisionados e não-supervisionados)
OBS: Caso você não possua uma base, utilize a base de dados de vinho brancos. Considere bons vinhos aqueles que obtiveram notas ≥ 6 . Como motivo da escolha (questão 2.a) apenas indique utilizou a base proposta pelo professor.
2. A partir da base escolhida, você irá aplicar as etapas do modelo [CRISP](#) de desenvolvimento de projetos em Ciência de Dados:
 - a. Explique a origem dos dados e o motivo para a escolha. Descreva, também, como os dados foram obtidos. Essa é a fase de **COMPREENSÃO DO NEGÓCIO**.
 - b. Descreva as variáveis do problema e o tipo de cada uma (categórica ou numérica). Mostre a distribuição (usando um histograma) para cada uma delas. Comente sobre a faixa dinâmica de cada uma delas. Essa é a fase de **COMPREENSÃO DOS DADOS**.
 - c. Descreva o objetivo do modelo que será criado neste projeto.
 - d. Agora faça o tratamento de dados necessários para o treinamento: escalonamento, normalização, transformação de variável (ex: aplicar a função log em uma variável com distribuição exponencial), separação entre treino e teste... Você deve optar pelas tarefas necessárias, descrevendo cada uma delas e justificando o motivo. Essa é a fase **PREPARAÇÃO DOS DADOS**.
 - e. Agora defina o modelo de redes neural do tipo MLP que você irá utilizar. Essa é a fase de **MODELAGEM**:
 - i. Quantas camadas?
 - ii. Quantos neurônios na camada de entrada?
 - iii. Quantos neurônios na camada de saída? Justifique.
 - iv. Quantos números de neurônios na camada intermediária? Como esse número foi escolhido?
 - v. Qual será o algoritmo utilizado no treinamento?
 - vi. Qual a função de ativação será utilizada em cada camada? Justifique.
 - vii. Qual a função de otimização será utilizada neste treinamento?
 - viii. Quais serão as figuras de mérito que serão usadas na avaliação do modelo. Justifique.
 - f. Apresente os seguintes resultados:
 - i. A matriz de confusão do problema

- ii. O(s) histograma(s) da(s) saída(s) da rede neural
 - iii. Os valores das figuras de métrica utilizadas.
 - iv. Caso tenha utilizado validação cruzada no treino, apresente as incertezas para cada um dos resultados anteriores.
- g. Avalie os resultados, dando sua interpretação de acordo com a compreensão da natureza do problema proposto. Essa é a fase de **AVALIAÇÃO DO MODELO**.

Assim que terminar, salve o seu arquivo HTML ou PDF (exportado de um Jupyter Notebook) e poste no Moodle. Utilize o seu nome para nomear o arquivo, identificando também a disciplina no seguinte formato: “nomedoaluno_nomedadisciplina_pd.html” ou “nomedoaluno_nomedadisciplina_pd.pdf”.