Para os resultados, iremos realizar os seguintes passos (com e sem a variável diff\_spd):

1. Dividimos os dados em treino/teste aleatoriamente numa proporção 70/30 (verificar balanceamento no conjunto treino e teste);

2. Faz-se a seleção de covariáveis para a Árvore de Decisão e Regressão Logística via Lasso com o conjunto de treino;

3. Dividem-se os dados de treino, novamente, aleatoriamente em treino e validação numa proporção 60/40;

4. Para cada um dos métodos (fora a árvore de decisão), fazemos o seguinte:

a. Bayes Ingênuo Gaussiano: ajusta-se, no conjunto de treino, modelo completo e com covariáveis selecionadas pela árvore de decisão e por lasso (passo 2), escolhendo aquele que apresentar melhor desempenho (Curva ROC "melhor", acurácia maior, sensibilidade maior, especificidade maior, etc.);

b. Bayes Ingênuo Flexível: ajusta-se, no conjunto de treino, modelo completo e com covariáveis selecionadas pela árvore de decisão e por lasso (passo 2), para cada um dos kernels, escolhendo aquele que apresentar melhor desempenho (Curva ROC "melhor", acurácia maior, sensibilidade maior, especificidade maior, etc.);

c. Regressão Logística: comparamos modelo com todas as covariáveis e modelo ajustado via lasso, escolhendo aquele que apresentar melhor desempenho (Curva ROC "melhor", acurácia maior, sensibilidade maior, especificidade maior, etc.). No caso do modelo via lasso apresentar melhor desempenho, guardamos o valor de lambda obtido.

5. Ao final do processo, calculam-se as médias de todos os resultados (métricas avaliativas e Curva ROC);

6. Apresentam-se os resultados e o melhor modelo é escolhido (Curva ROC "melhor", acurácia maior, sensibilidade maior, especificidade maior, etc.).

Seção de Resultados:

1. Comparação de modelos completos (Gaussiano x Melhor Flexível x Regressão Logística)
2. Seleção de Covariáveis
   1. Seleção Lasso e Seleção de Árvore (só vai dar velocidade, fazer comentário)

Usar modelo completo ou treino para selecionar covariáveis para nb?

Usar estimativa da área abaixo da curva ROC como métrica?