T2 Classical Techniques - Feedback of Mateus

Matheus Corrêa Lindino • 1:35 PM, May 29 (BRT)

Análise:

- Outliers: Sempre que possível, trate amostras outliers presentes no conjunto de dados. Essa etapa é importante para entender se esses valores atípicos podem afetar a qualidade e o desempenho do modelo.

Regressão Linear:

- Análise apenas no conjunto de treinamento: É essencial realizar qualquer investigação ou análise dos dados exclusivamente no conjunto de treinamento. Ao visualizar a distribuição do conjunto de teste, por exemplo, introduz-se um viés nos resultados.

- Análises com o mesmo conjunto de dados: Para futuras análises, é recomendável utilizar o mesmo conjunto de dados nas atividades para manter a consistência. Durante as atividades, o conjunto de dados varia entre utilizar todas as features, features com correlação diferente de zero e apenas as duas principais.

Regressão Logística:

- Análise apenas no conjunto de treinamento: É essencial realizar qualquer investigação ou análise dos dados exclusivamente no conjunto de treinamento. No cálculo ­do Mutual Information (MI) foi utilizado o conjunto completo, o que pode contaminar os resultados.

- GridSearch: A avaliação dos modelos está contaminada. A forma adequada de comparar os modelos do GridSearch seria apresentar a média dos folds do melhor modelo. No caso, foi realizada a predição no conjunto de treino.

- Matriz de confusão normalizada: Ao apresentar a matriz de confusão, é preferível exibir os dados normalizados, em vez de valores absolutos. Isso permite uma melhor interpretação das proporções e distribuição das predições em cada classe.

- Balanceamento de dados: Não é adequado balancear o conjunto de teste e validação, apenas o conjunto de treinamento.