

Lista 4 - Inteligência Artificial

Aluno

Nome: Kaiky França da Silva

Matrícula: 802169

Repositório: <https://github.com/GLKaiky/PucMinas>

Questão 1: Ajuste de Hiperparâmetros com BayesSearchCV

Para otimizar os hiperparâmetros dos modelos de aprendizado de máquina no problema do **Titanic**, utilizei o **BayesSearchCV**. Os testes foram aplicados ao **Random Forest** e à **Árvore de Decisão**, resultando nas seguintes acurácias:

- **Random Forest: 83%**
- **Árvore de Decisão: 82,02%**

Os resultados mostram que ambos os modelos apresentam desempenhos similares, com o **Random Forest** se destacando levemente por sua maior robustez.

Questão 2: Balanceamento da Base de Dados

Como a base de dados do Titanic é desbalanceada, testei diferentes métodos de balanceamento para equilibrar as classes. As abordagens analisadas foram **SMOTE**, **TomekLinks**, **RandomUnderSampler** e **ADASYN**. Os resultados obtidos foram:

- **SMOTE**
 - Precisão: **0,75**
 - Recall: **0,826**
 - F1-Score: **0,786**
- **TomekLinks**
 - Precisão: **0,731**
 - Recall: **0,826**
 - F1-Score: **0,776**
- **RandomUnderSampler**
 - Precisão: **0,674**
 - Recall: **0,870**
 - F1-Score: **0,759**
- **ADASYN**
 - Precisão: **0,740**
 - Recall: **0,826**

- F1-Score: **0,817**

O método **SMOTE** demonstrou o melhor desempenho global, conseguindo um bom equilíbrio entre precisão e recall. Já o **RandomUnderSampler** apresentou menor precisão, mas compensou com um recall mais alto.

Questão 3: Imputação de Dados Ausentes

Para lidar com valores ausentes, testei três métodos de imputação: **Média, Moda e KNNImputer**. Os resultados obtidos foram:

- **Média: 97,75%**
- **Moda: 97,19%**
- **KNNImputer: 97,19%**

A imputação utilizando a **média** apresentou o melhor desempenho. No entanto, os demais métodos também mostraram bons resultados, indicando que abordagens simples podem ser eficazes para esse tipo de problema.
