**Lista 4 – Inteligência Artificial – Joao Paulo Dias Estevao**

**Questão 1 - BayesSearchCV para ajuste de hiperparâmetros:**

**Ajuste de parâmetros e comparação de modelos**

[**Repositório Github**](https://github.com/joaodest/inteligencia-artificial/blob/master/Lista%202/DecisionTree_com_gridsearch.ipynb)

**Texto

O conteúdo gerado por IA pode estar incorreto.**

**Texto

O conteúdo gerado por IA pode estar incorreto.**

**Texto

O conteúdo gerado por IA pode estar incorreto.**

**Texto

O conteúdo gerado por IA pode estar incorreto.**

**Resultados para árvore de decisão:**

* **Melhores parâmetros**:
  + criterion: 'gini'
  + max\_depth: 7
  + max\_features: 'log2'
  + min\_samples\_leaf: 11
  + min\_samples\_split: 20
* **Métricas**:
  + Acurácia: 0.8380
  + F1-Score: 0.7820
  + Precision (classe 1): 0.90
  + Recall (classe 1): 0.69
* **Importância dos atributos**:
  + Sex (0.539060)
  + Pclass (0.135068)
  + Fare (0.114868)

**Resultados para Random Forest:**

* **Melhores parâmetros**:
  + criterion: 'gini'
  + max\_depth: 20
  + max\_features: 'sqrt'
  + min\_samples\_leaf: 1
  + min\_samples\_split: 20
  + n\_estimators: 300
* **Métricas**:
  + Acurácia: 0.8492
  + F1-Score: 0.8085
  + Precision (classe 1): 0.86
  + Recall (classe 1): 0.76
* **Importância dos atributos**:
  + Sex (0.331171)
  + Fare (0.164037)
  + Title (0.151659)

**Análise comparativa:**

1. **Desempenho dos modelos**:
   * O Random Forest obteve um desempenho superior em todas as métricas avaliadas (acurácia, F1-score, recall).
   * A diferença é mais significativa no Recall para a classe 1 (sobreviventes), indicando que o RF consegue identificar melhor os casos positivos.
2. **Importância dos atributos**:
   * Ambos os modelos identificaram "Sex" como o atributo mais importante
   * "Fare" (tarifa) apareceu entre os três mais importantes em ambos os modelos, sugerindo forte correlação entre o preço da passagem e a sobrevivência.
   * O Random Forest valorizou mais o "Title" (título/tratamento), enquanto a árvore de decisão deu mais peso à "Pclass" (classe da passagem).
   * Dos três atributos mais importantes, dois são comuns entre os modelos (Sex e Fare).
3. **Diferenças nos parâmetros**:
   * O Random Forest utilizou uma profundidade máxima maior (20 vs 7), mas compensou com um número maior de estimadores (300 árvores) e critérios de divisão mais estritos.
   * A estratégia para selecionar features difere: log2 para DT e sqrt para RF.

**Questão 2 - Balanceamento de Classes**

Para resolver esta questão, devemos aplicar diferentes técnicas de balanceamento à base de dados do Titanic, que apresenta desbalanceamento entre as classes de sobreviventes e não sobreviventes.

Para cada método, treinaremos o modelo de Random Forest (que teve melhor desempenho) com os mesmos hiperparâmetros otimizados e compararemos as métricas de F1-score.

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto. Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

**Questão 3 - Imputação de Dados Ausentes**

Para tratar os valores ausentes na base do Titanic, devemos comparar diferentes técnicas de imputação:

**Métodos:**

1. **Média/Mediana para variáveis numéricas**

**Texto

O conteúdo gerado por IA pode estar incorreto.**

**Calendário

O conteúdo gerado por IA pode estar incorreto.**

**Texto

O conteúdo gerado por IA pode estar incorreto.**

**Tela de celular com texto preto sobre fundo branco

O conteúdo gerado por IA pode estar incorreto.**

1. **KNNImputer**

**Texto

O conteúdo gerado por IA pode estar incorreto.**

**Tela de celular com texto preto sobre fundo branco

O conteúdo gerado por IA pode estar incorreto.**

1. **MissForest**

**Texto

O conteúdo gerado por IA pode estar incorreto.**