

Lista #4

Curso: Ciência da Computação

Disciplina: Inteligência Artificial

Prof^a. Cristiane Neri Nobre

Data de entrega: 23/03

Valor: 1,5 ponto

Questão 01

Utilizando o otimizador **BayesSearchCV** (from skopt import **BayesSearchCV**), ajuste os hiperparâmetros do Random Forest e Árvore de decisão para o problema do TITANIC.

Que modelo obteve o melhor desempenho? Quais os valores das métricas de avaliação?

Os atributos mais relevantes indicados pelo Random Forest e árvore de decisão são os mesmos? Discuta os resultados obtidos.

Questão 02

Uma vez que a base de dados do Titanic é desbalanceada, investigue métodos de balanceamento para balancear as classes. Discuta os resultados obtidos. Que método conseguiu ter um desempenho melhor de Precisão, Recall e F1-Score?

Para isto, veja o Slide “Parte 1 - **Processamento – Balanceamento**” que está no CANVAS.

- 1) Experimente pelo menos 3 métodos de balanceamento para balancear a base de dados da TITANIC e veja o que acontece com a qualidade da classificação.

```
from imblearn.over_sampling import SMOTE

from imblearn.under_sampling import TomekLinks

from imblearn.under_sampling import RandomUnderSampler
```

- 2) Experimente também o método **DSTO-GAN** que está em: <https://pypi.org/project/dsto-gan/>
Balanceie a base com este método e compare o resultado com os métodos anteriores.

Questão 03

Uma vez que a base de dados do Titanic possui dados ausentes, investigue métodos de imputação para imputar as ausências desta base de dados.

Para isto, veja o Slide “**Parte 2 - Processamento - Dados ausentes**” que está no CANVAS.

Experimente pelo menos dois métodos de imputação na base do TITANIC e veja o que acontece com a qualidade da classificação. Os melhores resultados são obtidos com qual método?

Investigue Média, Moda, MissForest KNNImputer, dentre outros.