



escola  
britânica de  
artes criativas  
& tecnologia

## Módulo | Análise de Dados: Fundamentos de Aprendizado de Máquina

Caderno de **Exercícios**

Professor [André Perez](#)

### Tópicos

1. Teoria;
2. Atributos categóricos;
3. Atributos numéricos;
4. Dados faltantes.

### Exercícios

#### 1. Pinguins

Neste exercício, vamos utilizar uma base de dados com informações sobre pinguins. A ideia é preparar a base de dados para prever a espécie do penguin (variável resposta) baseado em suas características físicas e geográficas (variáveis preditivas).

In [ ]:

```
import numpy as np
import pandas as pd
import seaborn as sns
```

In [ ]:

```
data = sns.load_dataset('penguins')
```

In [ ]:

```
data.head()
```

## 1.1. Valores nulos

A base de dados possui valores faltantes, utilize os conceitos da aula para trata-los.

```
In [ ]: # resposta da questão 1.1
```

## 1.2. Variáveis numéricas

Identifique as variáveis numéricas e crie uma nova coluna **padronizando** seus valores. A nova coluna deve ter o mesmo nome da coluna original acrescida de "\_std".

**Nota:** Você não deve tratar a variável resposta.

```
In [ ]: # resposta da questão 1.2
```

## 1.3. Variáveis categóricas

Identifique as variáveis categóricas nominais e ordinais, crie uma nova coluna aplicando a técnica correta de conversão a seus valores. A nova coluna deve ter o mesmo nome da coluna original acrescida de "\_nom" ou "\_ord".

**Nota:** Você não deve tratar a variável resposta.

```
In [ ]: # resposta da questão 1.3
```

## 1.4. Limpeza

Descarte as colunas originais e mantenha apenas a variável resposta e as variáveis preditivas com o sufixo "\_std", "\_nom" e "\_ord".

```
In [ ]: # resposta da questão 1.4
```

---