

# Lab. 2 - Análise Exploratória e Aprendizado Não-Supervisionado

CMC-13 Introdução a Ciência de Dados  
(Trabalho em Grupo de dois, três ou quatro alunos)  
Prof. Paulo André Castro

## 1. Objetivo

Exercitar e fixar conhecimentos adquiridos sobre Ciência de Dados, preparação de dados, análise exploratória de dados e Aprendizado Não-Supervisionado utilizando uma base de dados fornecida.

## 2. Descrição do Trabalho

### 2.1. Base de dados (dataset)

Existem 420 aves contidas neste conjunto de dados. Cada ave é representada por 10 medidas (características) que são comprimentos e diâmetro de cinco ossos do corpo da ave: Úmero (hum), Ulna(uln), Fêmur (fem), Tibiotarso (tib) e Tarsometatarso (tar).

Os comprimentos estão em centímetros e terminam com l, exceto pelo Tibiotarso que é apresentado em milímetros. As larguras estão em milímetros e terminam em w. Valores faltantes estão representados como string vazias.

### 2.2. Tarefas a Realizar

#### 1. Preparação dos Dados

Descrever os procedimentos necessários adotados para preparar os dados, para aplicar os métodos não-supervisionados

#### 2. Análise Exploratória

Plotar gráficos scatter-plot para as variáveis (*features*) numéricas do dataset.

Selecione ao acaso duas features e plote os histogramas destas e também da feature ID. Discuta se alguma delas aparenta ser uma distribuição normal ou uniforme.

#### 3. Aprendizado Não-supervisionado e clusterização

Agrupe os pontos de dados em 2, 3, 6 e 9 clusters utilizando o método K-means. Discuta qual seria o melhor agrupamento, e por quê.

## 3. Material a ser Entregue e Prazo

**Deve ser entregue um jupyter notebook (formato .ipynb) com descrição, comentários e código-fonte**

OBS: Entregar através do Google Classroom! Não compacte o arquivos em um zip (ou qq outro formato),!

A. Notebook com descrição, comentários e código-fonte (ver detalhes abaixo)

**Prazo de Entrega: 16/junho/2023;**

### Estrutura do Notebook

OBS:(arquivo em jupyter Notebook ) Intercalar tags de texto e código fonte observando estrutura indicada abaixo. Comentar o código nas célula de código

**Título: Lab. 2 - Análise Exploratória e Aprendizado Não-Supervisionado**

**Equipe: Nomes do membros da Equipe**

#### 1. Preparação dos Dados

Descrever procedimentos, resultados e gráficos (se for o caso) desta tarefa

Códigos relacionados a esta tarefa

#### 2. Análise Exploratória

Descrever procedimentos, resultados e gráficos (se for o caso) desta tarefa

Códigos relacionados a esta tarefa

#### 3. Aprendizado Não-supervisionado e clusterização

Descrever procedimentos, resultados e gráficos (se for o caso) desta tarefa

Códigos relacionados a esta tarefa

#### 4. Conclusões: Comentários e sugestões sobre o trabalho (complexidade/facilidade, sugestões, etc.).

Bom Trabalho!

Prof. Paulo André Castro