

Lab. 2 - Análise Exploratória e Aprendizado Não-Supervisionado

CMC-13 Introdução a Ciência de Dados

(Trabalho Individual ou em Grupo de dois ou três 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, e boxplot 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 porquê.

3. Material a ser Entregue e Prazo

Devem ser entregues um relatório e um notebook com o código

Entregar através do Google Classroom!

OBS: Não compacte os arquivos em um zip (ou qq outro formato), faça os uploads dos dois arquivos!

A. Relatório em formato pdf (ver detalhes abaixo)

B. Código em formato Notebook (ve detalhes abaixo)

Prazo de Entrega: 6/junho/2022;

Item A: Relatório

Estrutura do Relatório do Projeto (arquivo em formato pdf)

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

2. Análise Exploratória

Descrever procedimentos, resultados e gráficos desta tarefa

3. Aprendizado Não-supervisionado e clusterização

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

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

Item B: Código do Projeto

Código do Projeto (Formato jupyter notebook, Linguagem: Python , R ou Julia)

Siga a estrutura do relatório, para organizar o código no notebook

1. Preparação dos Dados

Códigos relacionados

2. Análise Exploratória

Códigos relacionados

3. Aprendizado Não-supervisionado e clusterização

Códigos relacionados

Obs:

Use células markdown antes de cada célula de código, para descrever qual o propósito do código em seguida

Bom Trabalho!
Prof. Paulo André Castro