



# Projeto

Neste projeto, espera-se que os alunos recolham dados e usem a linguagem Python para explorar dados, criar modelos e apresentar resultados. Este trabalho deve ser desenvolvido em notebooks do Jupyter. Deve ser escolhido um tema e os dados obtidos recorrendo a datasets abertos (por exemplo, Pordata, Comissão Europeia, Câmara de Lisboa, entre outros). Em seguida, uma (ou várias técnicas) deve ser aplicada. Exemplos de técnicas que podem ser usadas: regressão, regressão logística, random forest, análise de clusters, SNA, redes neuronais. De referir que alguns dos tópicos podem exigir mais auto-estudo do que outros.

As entregas incluem:

- Relatório (relatório deve ser enviado em ambos os formatos: .docx filetype e pdf filetype)
- Caderno Jupyter
- Conjunto de dados (s) (incluindo a fonte desses conjuntos de dados (s))
- Eventualmente, os estudantes também podem entregar resultados na forma de uma aplicação web desenvolvido em Flask

## 1. Relatório

O relatório deve ter a seguinte estrutura:

### 1. Introdução

Na introdução, os estudantes devem fornecer o contexto do projeto. Também é importante identificar o problema genérico que os alunos esperam resolver e o principal objetivo do trabalho empírico.

### 2. Revisão da literatura

O grupo deve identificar um pequeno grupo de artigos que possam ajudar a identificar trabalhos semelhantes desenvolvidos por outros autores.

### 3. Trabalho empírico

O trabalho empírico deve ser baseado num ciclo de vida da ciência de dados, da seguinte forma:

1. Contexto de dados
2. Recolha de dados
3. Preparação de dados
4. Exploração de dados

5. Modelização de dados

6. Avaliação

7. Apresentação de resultados

#### **4. Resultados e discussão**

Os resultados obtidos a partir do trabalho empírico devem ser comparados com resultados de outros autores da revisão da literatura.

#### **5. Conclusões**

Em que consistiu o propósito do trabalho? O que grupo concluiu da literatura e também do trabalho empírico?

#### **Referências**

Todas as referências do projeto devem estar no estilo APA v.6. É importante que os alunos usem um sistema de gestão de referência, como Zotero [<https://www.zotero.org/>]

Páginas entre 5 e 12 páginas. Deve ser seguido o formato apresentado.

## **2. Jupyter Notebook**

Todas as etapas de análise de dados devem ser claramente explicadas.

A equipa é incentivada a explorar outras técnicas diferentes das apresentadas em sala de aula. No entanto, os membros do grupo devem dominar as técnicas.

## **3. Conjuntos de dados**

Obter dados originais, interessantes e úteis é um trabalho árduo. Assim, é uma tarefa que é valorizada no projeto. As fontes de dados devem ser claramente identificadas. O processo de recolha de dados deve ser bem descrito.

A maneira mais fácil de obter conjuntos de dados é selecionar conjuntos de dados já usados em artigos publicados (por vezes publicados com os artigos). Obviamente, deve ser evitado se o grupo quiser ter uma boa nota. Mas é possível.

## **4. Web App Utilizando Flask**

Esta não é uma tarefa obrigatória.

## **5. Apresentação:**

11 de dezembro - apresentação final da classe. Todos os membros da equipa devem apresentar.