



Programação e Ciência de Dados

Projeto (2000/2021)

Neste projeto, espera-se que os estudantes explorem dados fornecidos, criando modelos e mostrando resultados. Devem utilizar como ponto de partida os dados fornecidos pelo professor, devendo ainda ser adicionados dados extra.

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. Deve indicar os objectivos específicos bem como a abordagem genérica que se espera seguir para resolver.

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 (POS-DS e CRISP-DM), 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 conclui da literatura e também do trabalho empírica?

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

Os dados base serão fornecidos, podendo ser:

- Dados relativos aos acessos ao ensino superior
- Dados de covid19

Os grupos escolherão o que entenderem melhor. Devem porém, pesquisar dados para enriquecer a análise.

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 descrita.

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:

15 e 18 de dezembro - apresentação final à turma. Todos os membros da equipa que vai apresentar deve estar no início da aula e permanecer até ao fim.

6. Cotações

1. Formato tem de ser seguido sob pena de não aceitação do trabalho (3 valores)
 - 1.1. Relatório (formato PDF e DOC/DOCX/ODT)
 - 1.2. Jupyter notebook (s)
 - 1.3. Dataset
 - 1.4. Aplicação para fazer entrega, exemplo Flask (opcional)
2. Jupyter (6 Valores):
 - 2.1. Compreensão do problema
 - 2.2. Recolha de dados
 - 2.3. Preparação de dados
 - 2.4. Modelos seleccionados
 - 2.5. Intepretação
3. Relatório (5 valores)
4. Apresentação em aula (5 valores; colegas contribuem com 75% da avaliação)
5. Avaliação dos grupos dos colegas (1 valor)