

# Mineração de dados

Ano letivo 2023/2024

## Enunciado dos trabalhos práticos

### I. Componente de trabalho em grupo:

- Trabalho em grupo a ser realizado por grupos de 3 ou 4 elementos; os grupos de trabalho deverão ser indicados até ao dia **1 de março** no google sheet disponível no link seguinte:

[https://docs.google.com/spreadsheets/d/1HqjXb-mqtSSrFbPTlkGGroGdruHmYJz6qI-VC\\_yyhaA/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1HqjXb-mqtSSrFbPTlkGGroGdruHmYJz6qI-VC_yyhaA/edit?usp=sharing)

(folha Grupos)

- **Âmbito do trabalho:** deverá constar da implementação de uma aplicação ou estudo que envolva a integração de:
  - um modelo de linguagem (LLM), que será escolhido pelo grupo mediante o seu desempenho no tema escolhido, recomendando-se o uso de LLMs em código aberto
  - o LLM será refinado/ estendido com a recolha, processamento e análise de dados de uma ou mais fontes publicamente acessíveis com o objetivo de melhorar/ estender o modelo anterior para aplicações específicas
  - o tópico do trabalho deverá ser definido pelo grupo até ao dia **22 de março** no link acima referido, indicando o objetivo da aplicação ou estudo a desenvolver, bem como as fontes de dados selecionadas; o docente irá colocando feedback em relação a estas escolhas
- O trabalho terá duas **apresentações** obrigatórias a realizar pelos elementos do grupo (todos deverão estar presentes) nas aulas dos dias **5 de abril** e **24 de maio**. Ambas terão a duração de 10 minutos. Todos os elementos do grupo deverão estar presentes em ambas.

- na 1ª apresentação, o grupo deverá apresentar objetivos do trabalho, as fontes de dados a usar, bem como a planificação do desenvolvimento do trabalho, incluindo as tarefas e a metodologia planeadas;
- na 2ª apresentação, o grupo deverá focar nos métodos e ferramentas usadas e desenvolvidas e nos resultados obtidos no trabalho; dado que depois da apresentação o grupo poderá ainda desenvolver o seu projeto, poderão ser indicadas tarefas ainda a realizar.
- O trabalho desenvolvido deverá ser disponibilizado num repositório github, cujo URL deverá ser colocado no google sheet acima referido até ao dia **12 de abril**, e onde deverão ser colocados todos os elementos relevantes para a avaliação:
  - slides das duas apresentações colocados nas datas das mesmas;
  - relatório final (deverá ser feito na forma de um **artigo** com um máximo de 10 páginas, excluindo referências);
  - código desenvolvido, dados e outro material suplementar relevante;
  - o artigo final e o restante conteúdo do repositório poderão ser atualizados até ao dia **14 de junho**
- A **avaliação** terá em conta as duas apresentações, bem como o artigo final e o código desenvolvido. Serão tidas em conta de forma integrada, nestes 3 elementos de avaliação: a originalidade da proposta e dos objetivos, a apresentação do trabalho relacionado, a complexidade técnica, a metodologia seguida, a qualidade da implementação, os resultados obtidos e a sua discussão.
  - Nas **apresentações** orais será valorizada a clareza na apresentação, a sua organização, a capacidade de discussão oral dos métodos e resultados e a capacidade de síntese/ cumprimento do tempo definido.
  - No **artigo** final, será valorizada a organização do documento, a correção e clareza da linguagem, a correta definição dos objetivos, a cobertura do trabalho relacionado, a relevância das referências bibliográficas, a explicação concisa e tecnicamente válida dos métodos e resultados, bem como a sua discussão.
  - Serão ainda tidos em conta todos os materiais suplementares no repositório que possam ajudar a aferir dos critérios acima referidos, nomeadamente o código usado, encorajando-se sempre que possível a utilização de exemplos documentados (e.g. notebooks).
  - O trabalho em grupo valerá **60% da nota final** da UC.

## II. Componente de trabalho individual:

- O trabalho individual constará do desenvolvimento de um “post” web (de tamanho sugerido entre 2 a 4 páginas impressas) sobre uma temática de interesse para a unidade curricular. Este tópico poderá ser a explicação breve de um tópico mais teórico, a resolução de um problema técnico específico ou um mini-tutorial de como abordar uma questão com software específico
- Os tópicos serão escolhidos pelos alunos, preenchendo o google sheet abaixo indicado até ao dia **12 de abril**;

[https://docs.google.com/spreadsheets/d/1HqjXb-mqtSSrFbPTlkGGroGdruHmYJz6qI-VC\\_yyhaA/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1HqjXb-mqtSSrFbPTlkGGroGdruHmYJz6qI-VC_yyhaA/edit?usp=sharing)

(folha Individual)

- O docente irá colocando feedback nas respetivas escolhas, sendo estas discutidas na aula desse mesmo dia, para esclarecimento de possíveis dúvidas.
- Os links para os trabalhos individuais de cada estudante serão disponibilizados num site web (na plataforma Notion). Cada site poderá ser desenvolvido também em Notion (recomendado) ou noutra tecnologia. Os links devem ser colocados no google sheet acima até ao final do dia **25 de abril** para todos os estudantes estarem em igualdade.
- As apresentações dos trabalhos serão realizadas nas aulas de **26 abril e 3 maio**; Cada apresentação deverá ter cerca de 5 minutos, consistindo num "pitch" que deve ilustrar o conteúdo do post e convencer à sua leitura completa.
- A avaliação terá em conta a avaliação do site pelo docente (considerando a sua originalidade e relevância, organização, qualidade técnica, linguagem), a sua apresentação e a avaliação dos pares (restantes estudantes da turma), valendo **40% da nota total da unidade curricular**.