



Universidade do Minho

Departamento de Informática

Mestrado [integrado] em Engenharia Informática

Dados e Aprendizagem Automática

1º/4º Ano, 1º Semestre

Ano letivo 2022/2023

Enunciado prático nº 2

6 de outubro de 2022

Tema *A Data Science Perspective*

Enunciado Uma exploração aprofundada dos dados permite que se tirem ilações, muitas vezes escondidas, que poderão ser importantes para se compreender o domínio e o problema em mãos. Com este enunciado prático é esperado que sejam aplicadas um conjunto de técnicas que permitam explorar e tratar *datasets*.

Tarefas Descarregar o *dataset* disponibilizado em <https://bit.ly/3ufnsgF> que contém dados de um conjunto de utilizadores de uma plataforma *web* assim como o seu sentimento em relação à mesma.

Realizar as tarefas seguintes:

T1. Carregar o *dataset*. Aplicar métodos para exploração de dados, i.e., analisar os dados em relação à sua:

- a. Tendência central;
- b. Dispersão estatística;
- c. Correlação entre *features*.

T2. Criar *plots* para visualização dos dados;

T3. Aplicar diferentes tratamentos de dados como por exemplo:

- a. Tratar valores em falta;
- b. Remover registos duplicados;
- c. Criar um atributo «*age_binned*» como resultado da criação de 3 *bins* de igual frequência sobre a *feature* «*age*»;
- d. Para cada registo, extrair o ano, mês e dia da semana da *feature* «*birthday*»;
- e. Remover utilizadores da plataforma que tenham, ao mesmo tempo, uma atividade na plataforma («*WebActivity*») inferior a 1 hora e que tenham mais de 70 anos;
- f. Por «*Sentiment Analysis*», obter a moda dos restantes atributos;
- g. Por «*Sentiment Analysis*» e «*MaritalStatus*», obter a média dos restantes atributos.

T4. Análise crítica sobre as tarefas anteriores. Que conclusões poderia uma empresa tirar?

Descarregar o *dataset* disponível em <https://bit.ly/3525yDr>. Este *dataset* contém dados referentes à performance de vários jogadores de futebol na edição 2017/2018 da *Premier League*.

Realizar a tarefa seguinte:

T5. Carregar o *dataset*. Explorar os dados, procurar informação relevante e mostrar essa mesma informação. P.ex., qual é a equipa mais indisciplinada? Qual é o top-10 dos assistentes para golo? Qual é o top-5 de nacionalidades nesta edição da *Premier League*?