



Universidade do Minho

Licenciatura em Engenharia Informática

Mestrado integrado em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º/4º Ano, 2º Semestre

Ano letivo 2021/2022

Enunciado Prático nº 3

Março, 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 os alunos deverão aplicar um conjunto de técnicas que permitam explorar e tratar *datasets*.

Tarefa 1 Numa primeira fase devem descarregar o *dataset users_sentiment* disponível na plataforma de e-learning da UMinho, secção [Conteúdo]. Este *dataset* contém dados de um conjunto de utilizadores de uma determinada plataforma *web* assim como o seu sentimento em relação à mesma.

T1. Carregar, no *Knime*, o *dataset Users_Sentiment* e aplicar nodos de exploração de dados como forma de permitir a análise dos dados em relação à sua:

- Tendência central;
- Dispersão estatística;
- Correlação entre *features*.

T2. Criar *plots* para visualização dos dados;

T3. Aplicar nodos para tratamento de dados de forma a:

- Excluir todas as colunas do tipo *Double*;
- Tratar valores em falta;
- Remover registos duplicados;
- Criar 3 *bins* de igual frequência para a *feature age*;
- Para cada registo, extrair o ano, mês e dia da semana da *feature birthday*;
- Excluir utilizadores da plataforma que tenham uma atividade na plataforma (*WebActivity*) inferior a 1 hora e que tenham mais de 70 anos;
- Excluir todos os registos que contenham a *sub-string* "co" no produto.

T4. Aplicar nodos para agregação de dados de forma a:

- Por género, obter o número e a percentagem de contratos, assim como a média da idade e da atividade na plataforma. Obter também o mínimo e máximo da idade;
- Por género e atividade na plataforma, obter a moda da análise do sentimento em relação à plataforma e a média da avaliação do sentimento;
- Por análise de sentimento, obter o número de contratos, a média do salário anual estimado, o somatório do salário anual e a média do número de contratos.

T5. Análise crítica à informação extraída das agregações efetuadas na tarefa anterior. Que conclusões poderia a empresa adquirir?

Tarefa 2

Numa segunda fase devem descarregar o *dataset EPL_Player_Stats* disponível na plataforma de e-learning da UMinho, secção [Conteúdo]. Este *dataset* contém dados referentes à performance de vários jogadores de futebol na edição 2017/2018 da *Premier League*.

T6. Carregar, no *Knime*, o *dataset EPL_Player_Stats* e proceder à exploração de dados, como forma de procurar e encontrar conhecimento relevante.

(e.g., qual a equipa mais indisciplinada? Qual o top-10 dos assistentes para golo? Qual o top-5 de nacionalidades na liga?)