



**Universidade do Minho**

Licenciatura em Engenharia Informática

Mestrado [integrado] em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º/4º Ano, 2º Semestre

Ano letivo 2022/2023

Ficha prática nº 2

Fevereiro, 2023

**Tema**

Exploração de dados com KNIME.

**Objetivos de aprendizagem**

Com a realização desta ficha prática pretende-se que os alunos:

- Conheçam a plataforma de análise de dados KNIME;
- Experimentem tarefas de exploração e preparação de dados;

**Enunciado**

Descarregue o *dataset* disponível na plataforma de *e-learning* da Universidade do Minho.

Este *dataset* contém dados de um conjunto de utilizadores de uma plataforma *web* assim como o sentimento identificado em relação à mesma.

Realize as tarefas seguintes:

T1. Carregar o dataset «Users\_Sentiment» e aplicar nodos de exploração de dados como forma de permitir a análise dos dados em relação a:

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

T2. Criar *plots* para visualização dos dados;

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

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

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

- 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;
- b. Por género e atividade na plataforma, obter a moda da análise do sentimento em relação à plataforma e a média da avaliação do sentimento;
- c. 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. Da análise crítica à informação extraída das agregações efetuadas na tarefa anterior, que conclusões poderia a empresa obter desta análise dos dados?