



Universidade do Minho

Licenciatura em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º Ano, 2º Semestre

Ano letivo 2024/2025

Guião prático n.º 4

Fevereiro, 2025

Tema

Exploração avançada e preparação de dados com KNIME

Objetivos de aprendizagem

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

- Explore a plataforma de análise de dados KNIME;
- Experimentem tarefas de exploração avançada de dados;
- Realizem tarefas de preparação de dados;

Enunciado

Os dados dizem respeito a diversas ações de recolha de informação meteorológica, de que resultaram dois ficheiros de dados do ano de 2020 e um ficheiro de dados de 2021.

Realizar as tarefas seguintes:

1. Carregar os dados de 2020 dos ficheiros disponibilizados;
2. Aplicar nodos para juntar dados de diferentes *datasets*,
 - a. Transformar o campo *dt_iso* em *Date&Time*;
 - b. Extrair campos do atributo *Date&Time* significativos para o problema;
 - c. Juntar os dois *datasets* em uma única estrutura de dados;
3. Aplicar nodos de exploração e tratamento de dados:
 - a. Remover colunas que partilhem a mesma informação;
 - b. Verificar e tratar a existência de entradas duplicadas;
 - c. Verificar e tratar a existência de *missing values*;
 - d. Agrupar os dados numa periodicidade diária, aplicando diferentes técnicas de agregação de dados;
 - e. Renomear as *features* depois de agrupar os dados;
 - f. Criar 3 *bins* de igual frequência para a *feature* "temp";
 - g. Verificar a distribuição da temperatura tendo em conta a *feature weather_main* (realizar este procedimento para diferentes métricas meteorológicas presentes no *dataset*);
4. Usar um nodo «Concatenate» para juntar os dados de 2020 com os dados de 2021;
5. Usar o nodo «Java Snippet»:
 - a. Criar um atributo com informação do trimestre do ano;
 - b. Criar um atributo com informação da estação do ano.



Descrição do *dataset* METRICAS

ATRIBUTO	DESCRIÇÃO
dt_iso	data e hora do registo da observação
temp	valor da temperatura, em Celsius
dew_point	ponto de condensação da água
feels_like	sensação térmica, em Celsius
temp_min	temperatura mínima, em Celsius
temp_max	temperatura máxima, em Celsius
pressure	valor da pressão atmosférica ao nível do mar, em hPa
humidity	valor da humidade, em percentagem
wind_speed	velocidade do vento, em m/s
wind_deg	direção do vento, em graus
rain_1h	volume de chuva na última hora, em mm
clouds_all	nebulosidade, em percentagem

Descrição do *dataset* DESCRICAO

ATRIBUTO	DESCRIÇÃO
dt_iso	data e hora do registo da observação
weather_main	grupo de parâmetros meteorológicos (p.ex., Rain, Clouds, Clear, etc.)
weather_description	dentro de cada grupo de parâmetros, descrição da condição climática