



Universidade do Minho

Licenciatura em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º Ano, 2º Semestre

Ano letivo 2024/2025

Guião prático nº 6

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

O enunciado diz respeito a dados de estudantes de 2 escolas distintas. O *dataset* «student_grades» contém dados relativos ao desempenho escolar dos alunos destas escolas.

Realizar as tarefas seguintes:

1. Injetar os dados na plataforma KNIME;
2. Aplicar nodos para exploração de dados para:
 - a. Analisar os dados em relação às suas características e padrões;
 - b. Verificar, graficamente, a distribuição dos estudantes nas diferentes escolas;
 - c. Verificar, graficamente, a distribuição do emprego das mães dos estudantes;
 - d. Agrupar por escola e obter a idade média dos estudantes que a frequentam;
3. Aplicar nodos de exploração de dados que respondam às seguintes situações:
 - a. A coabitação (ou não) dos alunos com os pais afeta a sua classificação final (G3)?
 - b. O sexo do estudante afeta a classificação final?
 - c. O tempo que demora a chegar à escola afeta as classificações?
 - d. Como se comporta a classificação final em relação à idade?
 - e. Como se comporta a classificação final em relação aos tipos de empregos que ambos os pais dos estudantes desempenham? E em relação ao nível de educação de ambos?
 - f. Há alguma relação entre a classificação final e o facto de os estudantes quererem frequentar o ensino superior?
 - g. Qual a relação entre o número de chumbos e a classificação final? E entre o número de chumbos e o tempo semanal de estudo?
4. Explorar as outras *features* e retirar informações relevantes sobre o *dataset* (p.ex.: fazer e responder a perguntas como em 3.).



Descrição do *dataset* STUDENT_GRADES

ATRIBUTO	DESCRIÇÃO
School	escola do estudante
sex	género do estudante
age	idade do estudante
address	morada do estudante ('U' - urbano, 'R' - rural)
famsize	tamanho da família ('LE3' - menor ou igual que 3, 'GT3' - mais do que 3)
Pstatus	situação de coabitação dos pais ('T' – juntos; 'A' - separados)
Medu	nível de educação da mãe (0 - nenhum, 1 - primária, 2 - 3.º ciclo, 3 - ensino secundário, 4 - ensino superior)
Fedu	nível de educação do pai (0 - nenhum, 1 - primária, 2 - 3.º ciclo, 3 - ensino secundário, 4 - ensino superior)
Mjob	tipo de emprego da mãe
Fjob	tipo de emprego do pai
reason	razão da escolha
guardian	quem detém a guarda do aluno
traveltime	tempo de viagem para a escola (1 - <15min., 2 - 15 a 30min., 3 - 30min. a 1h, 4 - >1h)
studytime	tempo semanal de estudo (1 - <2h, 2 - 2 a 5h, 3 - 5 a 10h, 4 - >10h)
failures	número de chumbos
schoolsup	suporte extra educacional
famsup	suporte familiar
paid	explicações pagas
activities	atividades extracurriculares
nursery	frequentou a creche
higher	pretende frequentar o ensino superior
internet	acesso à internet em casa
romantic	numa relação amorosa
famrel	qualidade das relações familiares (de 1 - muito mau a 5 - muito bom)
freetime	tempo livre depois da escola (de 1 - muito baixo a 5 - muito alto)
goout	saídas com amigos (de 1 - muito baixo a 5 - muito alto)
Dalc	consumo de álcool à semana (de 1 - muito baixo a 5 - muito alto)
Walc	consumo de álcool ao fim de semana (de 1 - muito baixo a 5 - muito alto)
health	estado de saúde (de 1 - muito mau a 5 - muito bom)
absences	número de faltas
G1	classificação do 1.º período
G2	classificação do 2.º período
G3	classificação final