



Universidade do Minho

Licenciatura em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º Ano, 2º Semestre

Ano letivo 2024/2025

Guião prático nº 2

Fevereiro, 2025

Tema

Exploração de dados com KNIME

Objetivos de aprendizagem

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

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

Enunciado

Descarregue o *dataset* disponível na plataforma de *e-learning* da Universidade do Minho, que contém dados de um conjunto de utilizadores de uma plataforma *web* assim como o sentimento identificado em relação à sua utilização.

Realize as tarefas seguintes:

1. Aplicar nodos de exploração de dados de forma a permitir a análise dos dados em relação:
 - a. Tendências centrais;
 - b. Dispersão estatística;
 - c. Correlação entre *features*.
2. Criar *plots* de visualização gráfica dos dados;
3. Aplicar nodos para tratamento de dados:
 - a. Excluir todas as colunas do tipo *Double*;
 - b. Tratar valores em falta (*missing values*);
 - 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 idade superior a 70 anos;
 - g. Excluir todos os registos que contenham o prefixo “co” no produto.
4. Aplicar nodos para agregação de dados de forma a:
 - a. Por género, obter a quantidade e a percentagem de contratos, o valor médio da atividade na plataforma (*WebActivity*), e, ainda, a idade média, mínima e máxima;
 - b. Por género e atividade na plataforma (*WebActivity*), obter a moda da análise de sentimentos em relação à plataforma e a média da avaliação de sentimentos;
 - c. Por cada tipo de sentimento (*SentimentAnalysis*), obter o número de contratos, a média do salário anual estimado, o somatório do salário anual e a média do número de contratos.
5. Utilizar técnicas de previsão:
 - a. Prever a idade (*string*) com os nodos DECISION TREE LEARNER/DECISION TREE PREDICTOR;
 - b. Prever a idade (*integer*) com os nodos LINEAR REGRESSION LEARNER/REGRESSION PREDICTOR;
 - c. Analise os resultados das previsões com os nodos SCORER/NUMERIC SCORER.

Descrição do *dataset* SENTIMENTS

ATRIBUTO	DESCRIÇÃO
Customer Key	Identificador do utilizador
Web Activity	Código identificador da atividade desenvolvida na plataforma <i>web</i>
Sentiment Analysis	Tipo de sentimento identificado no utilizador (<i>string</i>)
Sentiment Rating	Tipo de sentimento identificado no utilizador (código)
Marital Status	Estado civil S = Single (solteiro), M = Married (casado)
Gender	Sexo M = Male (masculino), F = Female (feminino)
Estimated Yearly Income	Vencimento anual estimado
Number of Contracts	Número de contratos
Age	Idade
Target	Meta alcançada (Y/N)
Available401K	Plano de poupanças (Y/N)
Customer Value Segment	Estrato económico
Churn Score	Probabilidade de abandono
Call Activity	Registo de chamadas
Products	Produtos detidos
birthday	Data de nascimento