



Curso: Mestrado Integrado em Engenharia Informática
U.C.: Descoberta do Conhecimento

Ficha de Exercícios 06	
Docente:	Hugo Peixoto José Machado
Tema:	RapidMiner - Regras de associação
Turma:	PL
Ano Letivo:	2020-2021 – 2º Semestre
Duração da aula:	2 horas

1. Parte I

- [1] O que são regras de associação? Para que servem?
- [2] Quais são as duas principais métricas calculadas nas regras de associação e como são calculadas?
- [3] De que tipo de dados devem ser os atributos de um dataset para usar os operadores Frequent Pattern no RapidMiner?
- [4] Como é que os resultados das regras de associação são interpretados? No exemplo dos slides desta aula, qual foi a regra mais forte e como se sabe?

2. Parte II

- [1] Faça download do dataset order.csv, importe-o para o RapidMiner e arraste-o para a janela de processo. Proceda à etapa de Data Understanding.
- [2] Conforme necessário, execute os passos referentes à etapa de Data Preparation no seu dataset. Certifique-se de que todas as suas variáveis tenham dados consistentes e que os seus tipos de dados sejam apropriados para o operador FP-Growth.
- [3] Gere regras de associação para o dataset. Modifique os valores de confiança (min confidence) e suporte (min support) para identificar os níveis ideais, de modo a obter regras interessantes com valores de confiança e suporte razoáveis. Analise as outras medidas de força das regras, como LaPlace ou Conviction. Documente as suas descobertas. Que regras encontrou? Que atributos estão mais fortemente associados? Existem produtos frequentemente conectados que o surpreendam? Quantas vezes tentou diferentes valores de suporte e confiança antes de encontrar algumas regras de associação? Alguma das suas regras de associação é boa o suficiente ao ponto de se basear nela para tomar decisões? Porquê?



[4] Crie um novo modelo de regras de associação usando o mesmo dataset, mas desta vez, use o operador WFPGrowth no RapidMiner. Para poder utilizar este operador, instale primeiro a extensão “Weka Extension” em Extensions -> Marketplace (procure Weka).

(Dicas para usar o operador W-FPGrowth: (1) Este operador cria as suas próprias regras sem a ajuda de outros operadores; e (2) Os parâmetros de suporte e confiança deste operador são identificados como U e C, respetivamente). Apresente e discuta os resultados obtidos.

[5] O algoritmo Apriori é frequentemente usado no processo de Data Mining para associações. Pesquise Apriori (W-Apriori) nos operadores do RapidMiner e adicione-o ao seu dataset num novo processo. Use o separador de Ajuda no canto inferior direito do RapidMiner para aprender sobre os parâmetros e funções desse operador. Apresente e discuta os resultados obtidos.

[6] Apresente uma conclusão global dos exercícios realizados, comparando os resultados obtidos através de cada uma das técnicas utilizadas