



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

Ficha de Exercícios 05	
Docente:	Hugo Peixoto José Machado
Tema:	Introdução ao RapidMiner
Turma:	PL
Ano Letivo:	2020-2021 – 2º Semestre
Duração da aula:	2 horas

1. Parte I

- [1] Quais as principais limitações de modelos de correlações?
- [2] O que é um coeficiente de correlação e como é interpretado?
- [3] Qual a diferença entre uma correlação negativa e uma correlação positiva?
 - a) Se dois atributos diminuem essencialmente à mesma taxa é uma correlação positiva e negativa? Explique.
- [4] Como é medida a força de uma correlação? Quais os limites para essa força?
- [5] Consegue pensar em atributos que poderiam ser interessantes incluir no dataset estudado no exemplo da aula?

2. Parte II

- [1] Aceda ao ficheiro *mpg_dataset.csv*.
- [2] Execute a operação de Data Understanding tendo em conta que:
 - cylinders: número de cilindros no motor
 - displacement: cilindrada do motor
 - horsepower: potência do motor
 - weight: peso do veículo
 - acceleration: aceleração do veículo, i.e., tempo em segundos para acelerar de 0 a 60
 - model year: ano do modelo do veículo nos anos 1900s
 - origin: origem do carro (1: Americana, 2: Europeia, 3: Japonesa)
 - mpg: consumo/eficiência de combustível (miles per gallon - mpg)



[3] Execute a etapa de Data Preparation no Weka ou no RapidMiner.

- Não se esqueça de analisar a existência de 'outliers' e 'missing values'. Depois de devidamente processados, guarde os dados num ficheiro .csv que permita a execução no rapidminer do processo de correlação.
- Importe o csv para o RapidMiner.
- Acrescente o csv no design.

[4] Documente quais os atributos que podem influenciar ou explicar o consumo/eficiência de combustível num determinado veículo (mpg).