

**Universidade do Minho**

Departamento de Informática

Mestrado Integrado em Engenharia Informática

Perfil de Sistemas Inteligentes

Unidade Curricular de Aprendizagem e Extração de Conhecimento

2015/2016

Aprendizagem e Extração de Conhecimento

André Geraldes a67673

Bruno Barbosa a67646

Francisco Dourado pg30458

**Abstract**

This report describes all the knowledge extraction procedures used in three distinct datasets. Every decision made throughout the project is presented as well as the relevant implementation details explored.

**Resumo**

Neste relatório é descrito todos os procedimentos de extração de conhecimento usados em três conjuntos de dados distintos. São apresentadas todas as decisões tomadas e explorados durante todo o processo assim como os detalhes de implementação mais relevantes.

Índice

[1 Introdução 4](#_Toc437555921)

[1.1 Enquadramento e objetivos 4](#_Toc437555922)

[1.2 Planeamento 5](#_Toc437555923)

[1.3 Estrutura do documento 5](#_Toc437555924)

[2 Desenvolvimento 6](#_Toc437555925)

[2.1 Dataset de Avaliação de Automóveis 6](#_Toc437555926)

[2.2 Dataset do Consumo de Energia Elétrica 6](#_Toc437555927)

[2.3 Dataset da Central de Energia Elétrica 6](#_Toc437555928)

[3 Anexos 7](#_Toc437555929)

[3.1 Exemplo 7](#_Toc437555930)

# Introdução

Os dados recolhidos a partir do mundo real não podem ser imediatamente utilizados para a extração de conhecimento visto que possuem imensos problemas. Problemas esses que podem estar relacionados com inconsistências, informação incompleta ou até mesmo lixo.

Surge, deste modo, a necessidade de tratar esses dados com a devida antecedência para que se adequem a uma determinada ferramenta de extração de conhecimento. A preparação dos dados para a extração de conhecimento segue um conjunto de passos bem definidos: seleção, pré-processamento, transformação, mineração dos dados (*Data Mining*) e interpretação.

Discretização, limpeza, integração, transformação e redução, são exemplos de tarefas indispensáveis desde que selecionamos os dados até à extração de conhecimento a partir dos mesmos.

## Enquadramento e objetivos

Com a liberdade de escolha dos conjuntos de dados a estudar, o nosso grupo focou-se em três *datasets* distintos. Um primeiro relacionado com a avaliação de automóveis, um outro associado ao consumo de energia elétrica e um terceiro relativo à energia gerada numa central elétrica. A escolha destes *datasets* teve em conta algumas considerações que achamos importante referir. O número de atributos deverá estar compreendido entre 3 a 6 unidades e o número de instâncias não deverá ser superior a 10000 nem inferior a 1500 unidades. A razão desta escolha deve-se ao facto de querermos trabalhar com *datasets* minimamente robustos mas que simultaneamente não sejam demasiado grandes.

Os objetivos deste trabalho consistem assim, na análise, interpretação e transformação dos dados e na extração de conhecimento através dos diversos relacionamentos entre os atributos.

## Planeamento

Cada elemento do grupo ficou responsável por tratar de um *dataset*. Neste caso, a distribuição foi feita da seguinte forma:

* André Geraldes – Avaliação de automóveis
* Bruno Barbosa – Consumo de energia elétrica
* Francisco Dourado – Energia gerada por uma central elétrica

## Estrutura do documento

O presente documento está dividido em dois capítulos e um anexo. No primeiro capítulo é feita uma breve exposição do problema descrevendo as considerações iniciais decididas pelo grupo.

No segundo capítulo, encontra-se toda a parte referente ao tratamento dos dados, que inclui uma exposição da cada *dataset* e todos os procedimentos aplicados sobre eles.

Nos anexos estão expostas algumas imagens que ajudam a complementar o trabalho desenvolvido.

# Desenvolvimento

Neste capítulo são apresentados todos os procedimentos utilizados durante a análise e o tratamento dos vários *datasets*, bem como as questões mais relevantes a eles associadas.

## Dataset de Avaliação de Automóveis

## Dataset do Consumo de Energia Elétrica

## Dataset da Central de Energia Elétrica

# Anexos

## Exemplo