



Universidade do Minho
Departamento de Informática

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

Enunciado do Trabalho Prático	
Docentes	José Machado e Cristiana Neto
Ano Letivo	2019-20 – 2º Semestre
Data de Apresentação	19 de junho de 2020 / 26 de junho de 2020
Data de Entrega	30 de junho de 2020

1. Tema

Com a realização deste trabalho prático pretende-se motivar os alunos, do Mestrado em Engenharia Informática, abrangidos pela unidade curricular de Descoberta de Conhecimento, para a análise e resolução de problemas com técnicas de Data Mining, em particular pela utilização das ferramentas lecionadas ao longo das aulas. Para tal, serão atribuídos diferentes *datasets* aos diferentes grupos, sendo que cada *dataset* representa um problema diferente. Concretamente, pretende-se que seja efetuada cada uma das fases do processo de Data Mining seguindo a metodologia CRISP-DM.

2. Estrutura

A componente prática da unidade curricular de Descoberta de Conhecimento consistirá na realização de um trabalho de grupo relacionado com as temáticas abordadas nas aulas teóricas e práticas. O trabalho deverá ser realizado de acordo com os objetivos propostos para o mesmo, e elaborado um artigo científico com a descrição e a explicação das tarefas desenvolvidas. O artigo científico deve ser redigido preferencialmente em inglês, de modo a potencializar a sua publicação. O limite do número de páginas do artigo científico é entre 10 a 15 páginas. Este artigo deve seguir o *template* fornecido pelos docentes.

3. Objetivos

O presente trabalho tem como objetivo avaliar os conhecimentos adquiridos nas aulas da unidade curricular de “Descoberta do Conhecimento”, nomeadamente no que diz respeito à análise e resolução problemas recorrendo a técnicas de *Data Mining*.

Em termos práticos o trabalho envolverá a utilização das seguintes ferramentas:

- Ferramenta que permita a aplicação de algoritmos de Machine Learning (Weka, RapidMiner, R);
- Editor de texto para a escrita do artigo científico.

A resolução dos problemas deve desenrolar-se segundo as fases da metodologia CRISP-DM, tal como abordado nas aulas:

- *Business Understanding;*
- *Data Understanding;*
- *Data Preparation;*
- *Modeling;*
- *Evaluation;*
- *Deployment.*

4. Avaliação

O trabalho prático será realizado em grupo. Cada grupo irá receber o seu dataset e deve responder a confirmar o recebimento do mesmo e a confirmar os elementos do grupo.

A sessão de apresentação dos trabalhos práticos decorrerá no dia 19 de junho. Caso não seja possível apresentar todos os trabalhos nesse dia, as apresentações serão concluídas no dia 26 de junho. A entrega final do trabalho compreende a entrega do artigo (formato pdf), que deve ser submetido na *blackboard*, dentro dos prazos e nos termos estabelecidos. A data limite para a entrega do trabalho prático é até às 24h00 do dia 30 de junho de 2020.

A avaliação final do trabalho será realizada, fundamentalmente, da seguinte forma:

- Avaliação das secções individuais do artigo (abstract, introduction, background, research methodology, results,...);
- Avaliação da escrita e estrutura do artigo;
- Avaliação da congruência do trabalho à publicação científica;
- Avaliação da utilização das ferramentas de *Data Mining*;
- Avaliação das apresentações semanais;
- Avaliação da apresentação final.