



Universidade do Minho
Departamento de Informática
Mestrado [integrado] em Engenharia Informática

Perfil de Machine Learning: Fundamentos e Aplicações
Sistemas Baseados em Similaridade
1º/4º Ano, 1º Semestre
Ano letivo 2019/2020

Trabalho Prático - 2ª Parte
Outubro, 2019

| | |
|----------------------------------|--|
| Tema | Conceção e implementação de modelos de <i>Machine Learning</i> usando Árvores de Decisão. |
| Objetivos de Aprendizagem | Com a realização deste trabalho prático pretende-se sensibilizar e motivar os alunos para a conceção e desenvolvimento de um projeto de <i>Machine Learning</i> utilizando a plataforma <i>Knime</i> e, como modelo de aprendizagem, Árvores de Decisão. |
| Enunciado | <p>A 2ª parte da componente prática de avaliação incidirá sobre as seguintes tarefas:</p> <ul style="list-style-type: none">• Os grupos deverão consultar, analisar e selecionar um <i>dataset</i> de entre os que estão acessíveis a partir de fontes como, por exemplo, o <i>Google Dataset Search</i>;• Para além do <i>dataset</i> selecionado no ponto anterior, os grupos deverão também trabalhar o <i>dataset</i> disponível em https://www.kaggle.com/c/tp2sbs2019:<ul style="list-style-type: none">○ O link anterior redireciona para a plataforma <i>Kaggle</i> onde foi criada uma competição. O <i>dataset</i> a utilizar na competição, assim como todos os detalhes do funcionamento da mesma, estão disponíveis no link acima referido. Em suma, deverão conceber e treinar um modelo, aplicá-lo e submeter os resultados na plataforma, a qual dará a <i>accuracy</i> do modelo desenvolvido;○ Na plataforma <i>Kaggle</i> deverão formar equipas com os restantes elementos do grupo de trabalho. O nome da equipa deverá seguir o formato GRUPO_X_MLFA onde X corresponde ao número do grupo;○ O link para participar na competição é o seguinte: https://www.kaggle.com/t/f6003551d2d449bf9602c3880c20de68• Utilizar a plataforma <i>Knime</i> para desenvolver um, ou vários, <i>workflows</i> para:<ul style="list-style-type: none">○ Exploração, análise e preparação dos dados de ambos os <i>datasets</i>;○ Extração de conhecimento dos dados;○ Utilizar Árvores de Decisão como modelo de <i>Machine Learning</i> para aprendizagem supervisionada;○ Obtenção de resultados, incluindo o <i>tuning</i> do modelo.• Interpretar os resultados e a sua utilidade no contexto dos problemas subjacentes aos <i>datasets</i> trabalhados. Determinar quais os resultados mais relevantes;• Criação de objetos visuais que permitam ter uma noção gráfica dos modelos e dos resultados obtidos; |

- Submeter os resultados obtidos no *Kaggle* de forma a obter a *accuracy* do modelo. Este passo refere-se, obviamente, apenas ao *dataset* da competição.

Os resultados obtidos deverão ser objeto de um relatório que contenha, entre outros:

- Quais os domínios a tratar, quais os objetivos e como se propõe a atingi-los;
- Qual a metodologia seguida e como foi aplicada;
- Descrição e análise detalhada de ambos os *datasets* e de todo e qualquer tratamento efetuado aos mesmos;
- Descrição dos *workflows* criados e com que objetivo. Quais os principais nodos e como foram configurados, entre outros detalhes que seja oportuno fornecer;
- Descrição detalhada do modelo desenvolvido, quais as suas características, como e sobre que parâmetros foi realizado o *tuning* do modelo, características do treino, entre outros detalhes que seja oportuno fornecer;
- Sumário dos resultados obtidos e respetiva análise crítica;
- Apresentação de sugestões e recomendações após análise dos resultados obtidos e dos modelos desenvolvidos.

Todo o processo deverá ser acompanhado de exemplos e indicações que permitam reproduzir todos os passos realizados assim como os resultados obtidos.

Entrega e Avaliação

As submissões deverão ser feitas por correio eletrónico para pjon@di.uminho.pt e para bruno.fernandes@algoritmi.uminho.pt, enviando, num único ficheiro compactado, todos os elementos produzidos. Tanto o assunto da mensagem como o ficheiro submetido deverão estar identificados na forma “SBS_TP2_MLFA_GX”, em que X designa o número do grupo de trabalho. As submissões deverão ser realizadas até ao final do dia 21 de novembro de 2019.

A sessão de apresentação do trabalho decorrerá no dia 21 de novembro de 2019, na sala DI-1.10, tendo início às 14h00min. Cada grupo disporá de 8 minutos para realizar a apresentação, utilizando os meios que considerar mais adequados.

A avaliação deste trabalho de grupo contará com os seguintes elementos:

- Pelo documento produzido, 75%;
- Pela apresentação realizada do trabalho desenvolvido, 25%.