



Universidade do Minho

Departamento de Informática

Mestrado [integrado] em Engenharia Informática

Mestrado em Engenharia de Sistemas

Perfil de Machine Learning: Fundamentos e Aplicações

Sistemas Baseados em Similaridade

4º/2º Ano, 1º Semestre

Ano letivo 2020/2021

Enunciado Prático nº 7 (Parte II)

10 de dezembro de 2020

<b>Tema</b>	<i>Tuning</i> de modelos de <i>clustering</i> , medidas de qualidade e criação de <i>datasets</i> usando <i>APIs</i> públicas
<b>Enunciado</b>	Pretende-se, com este enunciado prático, que sejam invocadas <i>APIs</i> públicas com vista à obtenção de dados para construção de <i>datasets</i> .
<b>Tarefas</b>	<p>Devem desenvolver um, ou mais, <i>workflows</i> para captura de dados referentes à qualidade do ar em diversas regiões a nível mundial, i.e.:</p> <p><b>T1.</b> Analisar a <i>Open AQ Platform</i>, a qual fornece uma <i>API</i> que disponibiliza dados sobre a qualidade do ar (ver <a href="https://docs.openaq.org/">https://docs.openaq.org/</a>). Efetuar um <i>GET request</i> que devolva todas as cidades portuguesas atualmente disponibilizadas pela <i>Open AQ</i>;</p> <p><b>T2.</b> Analisar e tratar a resposta JSON obtida em <i>T1</i>:</p> <ol style="list-style-type: none"><li>Passar o JSON para tabela, sendo que na configuração do nodo devem garantir (1.) a expansão dos <i>arrays</i> JSON para colunas e (2.) que o parâmetro <i>children expansion</i> vá, exatamente, até ao nível 2;</li><li>Remover todas as colunas exceto as dos resultados;</li><li>Obter a transposta e aplicar novamente o nodo <i>JSON to Table</i>.</li></ol> <p><b>T3.</b> Para cada cidade portuguesa disponibilizada pela <i>Open AQ</i>:</p> <ol style="list-style-type: none"><li>Obter os dados mais recentes sobre o nível de <b>ozono</b>, limitando o número de resultados obtidos a apenas 1 (ver <a href="https://docs.openaq.org/#api-Latest">https://docs.openaq.org/#api-Latest</a>);</li><li>Passar o JSON a tabela e filtrar todas as colunas exceto a <i>city</i>, <i>parameter</i>, <i>value</i> e <i>lastUpdated</i>. Fazer o <i>cast</i> da coluna <i>value</i> de <i>double</i> para inteiro;</li></ol> <p><b>T4.</b> Ordenar os registos de cada cidade por nível de ozono e implementar técnicas de visualização de dados numa vista web. Obter também dados referentes a outros parâmetros ambientais;</p> <p><b>T5.</b> Obter e visualizar dados de outras plataformas como a <i>OpenWeatherMaps</i> ou a <i>OpenUV</i>, por exemplo.</p> <p>Numa segunda fase, pretende-se, com este enunciado prático, que seja feito o <b>tuning de modelos de <i>clustering</i> utilizando o método do cotovelo</b>, i.e.:</p> <p><b>T6.</b> Consultar, selecionar e tratar um <i>dataset</i> sobre o qual deverá ser, numa tarefa posterior, aplicado um método de <i>clustering</i>;</p>

**T7.** De forma a tentar identificar, visualmente, *clusters* no *dataset* devem ser utilizados diagramas de dispersão. Deve também ser aplicada uma Análise de Componentes Principais (PCA) de forma a projetar os dados em apenas duas dimensões. O que se conclui?

**T8.** Aplicar o método do cotovelo de forma a identificar o número ótimo de clusters, utilizando *k-Medoids* como método de *clustering*. Utilizar a métrica *Mean Average Error* (MAE) como medida de qualidade;

**T9.** Desenvolver uma vista que permita a um utilizador analisar os gráficos gerados pelo método do cotovelo. Através de *widgets*, permitir que um utilizador defina o número de clusters a ser utilizado pelo método de *clustering*;

**T10.** Experimentar o método de *clustering k-Means* e comparar a sua performance com os resultados obtidos em T8.