



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 2

22 de outubro de 2020

Tema	<i>Churn Prediction</i>
Enunciado	<p>No sector das telecomunicações, <i>churn</i> é uma medida do número de clientes que estão a sair de uma operadora. Os clientes poderão estar de saída porque encontraram preços mais baixos na concorrência ou porque estão desagrados com o serviço prestado, entre outros motivos. Assim, para uma operadora de telecomunicações torna-se imperativo que existam modelos capazes de prever a possibilidade de <i>churn</i> de um cliente, isto é, a possibilidade de um cliente estar de saída. Isto permitirá que a operadora tente segurar o cliente antes que este opte pela saída, oferecendo melhores serviços ou preços mais atrativos.</p>
Tarefas	<p>Numa primeira fase devem descarregar dois <i>datasets</i> provenientes de uma operadora de telecomunicações. O primeiro (https://goo.gl/BSUhZ3) contém dados de chamadas de um cliente enquanto que o segundo contém dados contratuais (https://goo.gl/YZLDPf). Um valor de <i>churn</i>=0 significa que o cliente permaneceu na operadora; <i>churn</i>=1 representa clientes que abandonaram a operadora. Devem, de seguida:</p> <p>T1. Carregar, no <i>Knime</i>, ambos os <i>datasets</i>. Utilizar um nodo <i>Joiner</i> para agregar, por “<i>area code</i>” e “<i>phone</i>”, os dados provenientes das duas <i>readers</i>. Transformar o atributo <i>Churn</i> em nominal;</p> <p>T2. Particionar os dados de forma estratificada (pela <i>feature</i> “<i>Churn</i>”), utilizando 75% para aprendizagem e 25% para teste. Aplicar um <i>Decision Tree Learner</i> e um <i>Decision Tree Predictor</i>. Avaliar a precisão (<i>accuracy</i>) do modelo utilizando o nodo <i>Scorer</i> e a respetiva matriz de confusão;</p> <p>T3. Remover, iterativamente, <i>features</i> do <i>dataset</i> e reavaliar a performance dos modelos candidatos. Descrever os resultados obtidos;</p> <p>T4. Seguir as práticas de bons-hábitos na construção de <i>workflows</i>;</p> <p>T5. Utilizar o output de um nodo <i>Decision Tree Learner</i> para criar uma imagem de uma Árvore de Decisão e guardar essa imagem no ambiente de trabalho;</p> <p>T6. Explorar os <i>datasets</i> originais, procurando extrair informação relevante dos dados.</p>