



Universidade do Minho

Departamento de Informática

Mestrado [Integrado] em Engenharia Informática

Dados e Aprendizagem Automática

1º Ano, 1º Semestre

Ano letivo 2021/2022

Enunciado Prático nº 3

28 de outubro de 2021

Tema	Validação de Modelos e Métricas de Qualidade
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 desagradosos 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 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 os <i>datasets</i> e fazer <i>merge</i>, de ambos, por “Area Code” e “Phone”. Devem, de seguida, transformar o atributo <i>Churn</i> num atributo nominal;</p> <p>T2. Aplicar métodos para exploração e visualização de dados;</p> <p>T3. Usando uma Árvore de Decisão como classificador (<i>sklearn.tree.DecisionTreeClassifier</i>), avaliar a <i>accuracy</i> do modelo na previsão de <i>churn</i>. Avaliar também o modelo usando a métrica <i>f1_macro</i>. Utilizar <i>10-fold cross validation</i>:</p> <p style="padding-left: 40px;"><i>Nota:</i> Definir o X e o y. Atenção ao tipo dos atributos que fazem parte do X;</p> <p>T4. Obter matrizes de confusão do modelo e efetuar a respetiva análise crítica. Que conclusões se poderão tirar?</p> <p>T5. Alterar hiper-parâmetros da Árvore de Decisão (<i>criterion</i> e <i>max_depth</i>). Qual a variação na performance do modelo subjacente a estas alterações?</p>