



Universidade do Minho

Departamento de Informática

Mestrado [integrado] em Engenharia Informática

Dados e Aprendizagem Automática

1º/4º Ano, 1º Semestre

Ano letivo 2023/2024

Enunciado Prático nº 2

28 de setembro de 2023

Tema Validação de Modelos e Métricas de Qualidade

Enunciado No setor das telecomunicações, *churn* 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, como em qualquer outra área de negócios, torna-se imperativo que existam modelos capazes de prever a possibilidade de *churn* 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.

Tarefas Descarregar dois *datasets* contendo dados 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 *churn* = 0 significa que o cliente permaneceu na operadora; *churn* = 1 representa clientes que abandonaram a operadora.

Pretende-se:

T1. Carregar os *datasets* e fazer *merge*, de ambos, por “Area Code” e “Phone”. Devem, de seguida, transformar o atributo *Churn* num atributo nominal;

T2. Usando uma Árvore de Decisão como classificador (*sklearn.tree.DecisionTreeClassifier*), avaliar a *accuracy* do modelo na previsão de *churn*. Avaliar também o modelo usando a métrica *f1_macro*. Utilizar *10-fold cross validation*:

Nota: Definir o X e o y. Atenção ao tipo dos atributos que fazem parte do X;

T3. Obter matrizes de confusão do modelo e efetuar a respetiva análise crítica. Que conclusões se poderão tirar?

T4. Alterar hiperparâmetros da Árvore de Decisão (*criterion* e *max_depth*). Qual a variação na performance do modelo subjacente a estas alterações?