



**Universidade do Minho**

Licenciatura em Engenharia Informática

Mestrado integrado em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º/4º ano, 2º Semestre

Ano letivo 2021/2022

Enunciado Prático nº 2

Março, 2022

**Tema** *Churn Prediction*

**Enunciado** No setor das telecomunicações, *churn* é um termo usado para identificar uma métrica 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 (ou para qualquer outra entidade comercial ou industrial) torna-se imperativo que existam modelos capazes de prever a possibilidade de “rotatividade” dos clientes (*churn*), isto é, a possibilidade de um cliente estar de saída.

Esta vantagem permitirá ao operador segurar o cliente antes que este opte pela saída, oferecendo melhores serviços, preços mais atrativos, ou outros produtos ou ofertas mais competitivas.

**Tarefas** Deverão começar por descarregar os ficheiros disponíveis na plataforma de *e-learning* da UMinho, secção [Conteúdo]. O *dataset churn*, é um ficheiro compactado que contém 2 outros ficheiros: o ficheiro .XLS contém dados de chamadas de clientes, e o ficheiro .CSV contém dados com informação contratual. O valor de *churn=0* significa que o cliente permaneceu na operadora e o valor *churn=1* significa que o cliente abandonou a operadora.

Seguidamente, realizem as tarefas:

**T1.** Carregar, no *Knime*, ambos os *datasets*. Utilizar um nodo *Joiner* para agregar, por “*area code*” e “*phone*”, os dados provenientes dos dois *readers*. Transformar o atributo *Churn* em nominal;

**T2.** Aplicar nodos para exploração de dados, i.e., analisar os dados em relação às suas características e padrões, procurando extrair informação relevante dos dados;

**T3.** Particionar os dados de forma estratificada (pela *feature* “*Churn*”), utilizando 75% para aprendizagem e 25% para teste. Aplicar um *Decision Tree Learner* e um *Decision Tree Predictor*. Avaliar a precisão (*accuracy*) do modelo utilizando o nodo *Scorer* e a respetiva matriz de confusão;

**T4.** Remover, iterativamente, *features* do *dataset* e reavaliar o desempenho dos modelos candidatos. Descrever os resultados obtidos;

**T5.** Seguir as práticas de bons-hábitos na construção de *workflows*;

**T6.** Utilizar o output de um nodo *Decision Tree Learner* para criar uma imagem de uma Árvore de Decisão e guardar essa imagem no ambiente de trabalho.