



Universidade do Minho

Departamento de Informática

Mestrado em Matemática e Computação

Mestrado em Engenharia de Sistemas

Perfil de Machine Learning: Fundamentos e Aplicações

Sistemas Baseados em Semelhança

1º/2º Ano, 1º Semestre

Ano letivo 2023/2024

Enunciado Prático nº 2

28 de setembro de 2023

Tema

Churn Prediction

Enunciado

No sector 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 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

Numa primeira fase devem descarregar dois *datasets* provenientes de uma operadora de telecomunicações. O primeiro contém dados contratuais (<https://goo.gl/YZLDPf>) enquanto que o segundo contém dados de chamadas dos clientes (<https://goo.gl/BSUhZ3>). Valores de *churn*=0 significa que o cliente permaneceu na operadora; *churn*=1 representa clientes que abandonaram a operadora. Devem, de seguida:

T1. Carregar, no *Knime*, ambos os *datasets*. Utilizar um nodo *Joiner* para agregar, por “*area code*” e “*phone*”, os dados provenientes das duas *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 70% para aprendizagem e 30% para teste. Aplicar um *Decision Tree Learner* e um *Decision Tree Predictor*. Avaliar a precisão (*accuracy*) do modelo e a respetiva matriz de confusão;

T4. Remover, iterativamente, *features* do *dataset* e reavaliar a performance 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.