Universidade do Minho
Departamento de Informática

Dados e Aprendizagem Automática
Ano letivo 2025/2026

Practical Exercise no. 2

Theme *Quality Metrics and Model Validation*

Exercise

In the telecommunications sector, *churn* is a measure of the number of customers leaving an operator. Customers may be leave for various reasons, such as finding lower prices in the competition or being unhappy with the service provided. Therefore, as in any other business area, it is imperative for a telecom's operator to have models capable of predicting customer *churn*, i.e., the likelihood of a customer leaving. This will allow the operator to retain customers by offering better services or more attractive prices before they choose to leave.

Tasks

Download the two datasets containing data from a telecommunications operator. The first, *CallsData* (https://goo.gl/BSUhZ3), contains call data from a customer, and the second, *ContractData* (https://goo.gl/YZLDPf), contains a contract data.

A value of *churn = 0* means that the customer has remained with the operator, while a value of *churn = 1* means that the customer left the operator.

The aim is to:

T1. Load the datasets and merge them by "*Area Code*" and "*Phone*". Then transform the "*Churn"* attribute into a nominal attribute;

T2. Use a Decision Tree as a classifier (*sklearn.tree.DecisionTreeClassifier*) and evaluate the model's accuracy in predicting the *churn*. Also evaluate the model using the *f1_macro* metric. Use *10-fold cross validation*;

Note: define X and y., paying attention to the type of attributes that are part of X.

T3. Obtain the model's confusion matrices and analyse them critically. What conclusions can be drawn?

T4. Change the hyperparameters of the Decision Tree (*criterion* and *max_depth*). What variation in model performance is underlying these changes?