



**Universidade do Minho**

Departamento de Informática

Mestrado [Integrado] em Engenharia Informática

Mestrado em Matemática e Computação

Dados e Aprendizagem Automática

1º/4º Ano, 1º Semestre

Ano letivo 2024/2025

Practical Exercise no. 7

**Theme** Interpretability and Explainability

**Exercise** While interpretability focuses in understanding the decision-making process of a model, explainability focuses on explaining it to the user. Both are important because they provide insight into how decisions are made by machine learning algorithms, and essential to ensure that the models are reliable, secure, and adhere to ethical principles appropriate to the context.

The aim of this practical statement is to carry out a series of tasks that will give you a better understanding of each concept.

**Tasks** This practical exercise includes the following tasks:

**T1.** Follow and implement the steps presented in the lab notes in order to have a local and global interpretability of both features and models;

**T2.** Analysing the feature importance of the Titanic Dataset, which three features have higher decision weight? Analyse using Random Forest Feature Importance, Mean Decrease in Impurity and Permutation Importance;

**T2.** Analyse and compare the results obtained for Random Forest Classifier and Linear Regression using SHAP values. Which three features were more relevant for predicting the survival of a passenger in each model? How can it explain the model's decision?

**T3.** Conduct an analogue study using XGBoost. Compare the results obtained with the ones from the previous models. Which are the most important features to determine the target variable? Which model has higher interpretability/explainability?