

# Machine Learning and Data mining

## Serie 1. Simple rules

Prof. J. Savoy

October 3, 2022

1. In the folder “Data” you have access to the dataset Titanic.csv presenting information about travellers with their status (survived=1 (yes) or =0 (no)). In addition, you have the information about the class (Pclass), name (Name), gender (Sex), age (Age), sibling or spouse on board (1/0), parents or children aboard (1/0), and fare price (Fare).
  - (a) What is the best default rule for this dataset? (Default means without any evidence about the person)?
  - (b) What is the best 1R for this dataset?
  - (c) Can you produce a second rule based on a single attribute with a good effectiveness?  
You need to split the dataset into two disjoint sample, the training and the test set. For example, used 75% for the training sample, and the remaining 25% for the test set.
2. Using your select stock / market index and your decision criterion (binary or ternary) on the daily return of the next day or on the trend (daily return after 5/10 days), can you generate a 1R model using as possible predictor the volume, and the moving average (with a period of 5, 10, 20 50 or 200). You can learn on all days except the last 100 (that will be used as the test set).