

# SUPERVISED AND EXPERIENTIAL LEARNING

## (Master in Artificial Intelligence, UPC-URV-UB)

Spring semester, Course 2020/2021

March 12<sup>th</sup>, 2021

### **Practical Work 1 (PW1, Individual): a rule-based classifier**

The objective of this exercise is to implement and validate a *rule-based classifier*. Each student has a concrete rule-based classifier to implement according to this formula:

$$\text{NumCla} = \text{NumStud} \bmod 4$$

where,

*mod* is the remainder operator of the integer division

$\text{NumCla} = 0 \Rightarrow \text{Rule-Based Classifier} = \text{RULES}$

$\text{NumCla} = 1 \Rightarrow \text{Rule-Based Classifier} = \text{PRISM}$

$\text{NumCla} = 2 \Rightarrow \text{Rule-Based Classifier} = \text{CN2}$

$\text{NumCla} = 3 \Rightarrow \text{Rule-Based Classifier} = \text{RISE}$

The implemented classifier will be evaluated in several domains. The main steps that students must undertake are listed below.

The NumStud is provided in a separate file (NumStud-SEL-2021.pdf).

### **Procedure**

1. *Implement the corresponding rule-based classifier assigned to you in your selected programming language (Java, C++, R, Python, etc.).*
  - a. *The classifier must be able to read a supervised dataset in csv file format*
  - b. *Then, it should induce the set of rules from the training data set*
  - c. *The set of rules should be displayed in an interpretable way*
  - d. *The coverage and precision of each rule must be computed*
2. *Implement a rule interpreter being able to apply a set of classification rules to a validation/test dataset, obtaining the corresponding classification accuracy values.*
3. *Evaluate the set of rules obtained with the classifier, in at least 3 databases (one small, one medium and one large). You can use databases from UCI ML repository or other sources. Small  $\approx$  (# instances  $\leq 500$ ), Medium  $\approx$  ( $500 < \#$  instances  $\leq 2000$ ), and Large  $\approx$  (# instances  $> 2000$ ).*

### **Deliverable**

A ZIP file labelled as “PW1-SEL-2021-NameSurname”, delivered through “Racó de la FIB” (in the “Practical” tab) with the following content:

1. A folder named **“Documentation”** with a report (**maximum 20 pages on 11 pt. letter size**) containing:
  - a. Pseudo-code of *your implemented algorithm* of the rule-based classifier
  - b. Evaluation of results for all the tested databases:
    - i. Set of rules obtained with coverage and precision
    - ii. Comments on the interpretation of the rules
    - iii. Classification accuracy results
  - c. Precise instructions on how to execute the code
  - d. Other comments
2. A folder named **“Data”** with the files with the original dataset/s or database/s used both for training and for testing.
3. A folder named **“Source”** containing the source code of the implementation
4. An **executable object file** (\*.jar, etc.), if available
5. A **README.txt** file specifying the structure and contents of the ZIP file

Students must deliver the ZIP file on **9/4/2021**.

### **Qualification**

The qualification of this work will take into account the quality/functionality of the software delivered (correctness, efficiency and scalability), the robustness of the code, and the written documentation delivered.

**PW1 is due on April 9<sup>th</sup>, 2021**