

Data Stream Learning and Application

TP N° 1

Online Machine Learning Model using River

Submission date : 05/11/2023 à 22h00

The purpose of this lab is to practice data stream processing using the well known framework River. You will then perform different experiments on different classifiers using River and compare their performance giving guidance for the best choice.

What to do ?

This is a two parts lab. First, you will discover the framework River. Second, you will compare classifiers performance using River.

1. Implement an online learning with practical examples - Download the zip code of Professor Albert Bifet River Lab ([Source](#)) and follow the instructions described in the Jupyter Notebook. Discuss the results and comments the plots inside the Notebook.
2. Classifiers performance comparison
 - Download Covertypes dataset available [here](#)
 - Run several experiments with at least 5 classifiers, to find the classifier with better performance for this dataset
 - Write a short report presenting (Max 3 pages)
 - classifiers selected for the experiments, the results of the experiments, a discussion about the results and which classifier do you recommend to use with the Covertypes dataset.
3. Select [any dataset from River](#) (classification, regression, clustering), choose at least 3 methods from River API and compare their performance using different [metrics](#). Discuss the results in the same report.

What to submit ?

One compressed file (`.zip`, `.rar`, `.tar.gz`...) which will contain :

1. The Jupyter Notebooks for three parts. On top of each Jupyter Notebook, put a cell with the team members name and **'html' version of the notebooks**
2. The synthetic report discussing the results of part 2 and 3.

How to submit ? Do the 2 steps

1. **First, on the Moodle of Polytechnique** : Upload the compressed file and name it as follow: 'Student1LastName_Student2LastName_lab1.zip'
2. **Then, By email** : Send the two **HTML version of the notebookz + PDF report** (Maximum 5 pages) to mariam.barry@polytechnique.edu ⇒ EMAIL OBJECT SHOULD BE - IP Paris - M2DS - Lab 1 - River - FirstStudent FirstName_LASTName + Second_student_FirstName_LASTName. Do not include the PDF report inside the ZIP.