SUMMER SCHOOL 2

Welcome,

As a part of Data4Water EU Twining project, a series of three summer schools are organised. The second Summer School is organised by project partners from University of Milano-Bicocca and University Politehnica of Bucharest.

The Summer School topic is '***Using Data Science for Urban Water Management'*** and takes place from 12th June to 23rd June 2017. The participants of the Summer School will be offered a training on the same topic, organised on 31st May to 02nd June. Participants are strongly advised to participate at the training, as all the theoretical input will be offered during this training.

Detailed description of the content, logistics and venue of the school is given bellow.

Applications for the Summer Schools are welcomed, no later than 28th May, 2017. We look forward to seeing you in Bucharest soon.

Objective:

The program of the Summer School is designed to teach participants the necessary background on *Data Science* and *Machine Learning* for solving problems related to time series data and, in particular to urban water management. At the end of the program, participants will be able to solve their own water management related problems using mechanisms and tools from Data Science, addressing both batch and online Machine Learning tasks.

Description:

The Summer School schedule is:

1. Training courses – 2 ½ days [31st May to 02nd June].
2. Assignment of tasks for the participants – 1/2 day [02nd June]
3. Problem solving sessions guided by experts – 10 days [12th June to 22th June]
4. Session for presenting the final results - 1 day [23th June]

The detailed syllabus of the training courses covers two main aspects of data science field, as follows:

Topic 1 – Time Series Data Analysis

- Nature of the time series data
- Time Series data representation (raw data, features extraction, modelling)
- Statistical inference vs Machine Learning in the Time Series analysis domain
- Main Time Series analysis tasks (Indexing, Clustering, Classification, Prediction, Summarization, Anomaly Detection, Segmentation)
- Evaluating Similarity among time series (similarity in time, shape and change)
    - Factors influencing similarity (noise, drift, scaling, discontinuities, translation)
    - Cosine Similarity vs Dynamic Time Warping
- Time Series Clustering
- Time Series Prediction/Forecasting
- Hyperparameters tuning / algorithmic configuration through Sequential Model Based Optimization
    - Design variables

- o Performance measures as objective function
  - o Bayesian Optimization for Machine Learning
- An example: time series clustering and (SVM) regression for urban water demand forecasting in the short term (24-hours)

Topic 2 – Online Machine Learning for Fast/Streaming Data
- Big versus Fast/Streaming data
- Online Machine Learning challenges: learning under concept drift/shift
- Characteristics and differences with respect to "traditional" Machine Learning
- Online Machine Learning schemes: Incremental (online) vs two-phase (online/offline or micro-batch) learning
- Algorithms for online- clustering, classification, regression, anomaly detection, drift/shift detection
- Evaluation procedures in the online setting
- Some available tools for online machine learning
- An example: urban water demand forecasting through online machine learning

Content from all above two aspects will be addressed during the summer school.

Basic knowledge is required on the following topics:
- Statistics for time series analysis (i.e. AR, MA, ARMA, ARIMA, S-ARIMA models)
- Machine Learning and Data Mining
- Some specific learning strategies: k-means for clustering; Support Vector Machine and Artificial Neural Networks for regression
- R language

Application procedure:
To apply for a place please send a short CV (including research interests) to elena.apostol@cs.pub.ro or catalin.leordeanu@cs.pub.ro.
The Summer School is targeted at PhD/MD/undergrads students and other early career researchers. There is **no** participation **fee**.

Important dates:
The closing date for applications is: 28th May 2017
The final list of participants and detailed schedule will be announced on 29th May 2017
Please direct all enquiries to elena.apostol@cs.pub.ro or catalin.leordeanu@cs.pub.ro.

Organisers:
University of Milano-Bicocca: Antonio Candelieri and Francesco Archetti
University Politehnica Bucharest: Mariana Mocanu, Elena Apostol and Catalin Leordeanu