

## **Linking data mining, spatial analysis and algorithmic design: a review on a primer workshop based on Python**

*João V. Lopes*

The field of data mining, the practical application of machine learning, has recently become a full flagged science known as Data Science. An interdisciplinary discipline in the intersection of A.I., computer science, statistics, data visualization and database management, its main objectives are pattern recognition/knowledge discovery in datasets and prediction/data modelling. The application of this latter objective has become recently the subject of intense debate following the case of its use outside the scientific research. From political campaigns to the first mortal accident involving a self-driven car, these events brought the field to the highlights and, although its tools are not new, the scale of their implementation raises important questions considering their application, the nature of personal digital data and free will.

In research its application is most relevant in data rich fields and quantitative analysis. Here we can include spatial and urban analysis, which nowadays deal with huge datasets, e.g. combining Big Data from the internet, time series or unstructured data fluxes with urban form and structure, helping to assess or construct new investigation hypotheses. Using case-based reasoning and optimization data mining becomes a predictive tool able to assist the design process, producing scenarios or helping to explore constrained design solution spaces. The presentation will (i) briefly introduce the topic of data mining; (ii) its usage in urban analysis and design, and, mainly, (iii) report on a preliminary evaluation of the related workshop carried out in the context of the present seminar. The workshop introduces data mining to participants in a hands-on approach, focusing in simple tasks so concepts are internalized by playing with tools and scripting. The focus is on python scripting using Anaconda python data analysis package and Jupyter interactive Notebooks. In this way participants get a glimpse on one of the most flexible and widely used programming languages across a variety of fields, from algorithm design to data analysis, that is able of customize the tools that sometimes customize our own investigation or practice.