# PRACTICAL APPLICATION
## *MACHINE LEARNING*

The works must be structured as a scientific paper with sections like: Introduction, Problem description, Methodology, Results, Discussion, Conclusion, References. Interpretation of the models is fundamental, taking into account the variables involved.

Practical application 1 (classification): Using *Weka* software is recommended.

Chose a dataset (public or from the student himself). The dataset should be sized to at least 15 variables and 100 observations.

The **supervised** classification algorithms to be used are all seen in class, both probabilistic and non-probabilistic. There will be three analyses: (1) with all original variables; (2) with a univariate filter feature subset selection; (3) with a multivariate (filter and wrapper) feature subset selection. Metaclassifiers will also be used. All merit figures should be estimated with an honest method.
Also, all the **unsupervised** classification algorithms seen in class will be used. The class variable obviously will be excluded.

Practical application 2 (probabilistic graphical models): Choose a dataset and learn a Bayesian network and show and discuss the relationships found. Using *BayesFusion (https://www.bayesfusion.com/)* is recommended. Perform inferences (exact and approximate) with the model.

Practical application 3 (spatial statistics): Choose a dataset with spatial characteristics and follow the steps seen in class. Using the *spatstat* R package is recommended.

OBSERVATION: The maximum number of pages allowed are: 20 (Practical appl. 1), 15 (Practical appl. 2) and 10 (Practical appl. 3), with a font size of 11 or 12pt.

The document will be sent via Moodle. Otherwise it will not be considered and the student will fail this part of the subject. Please consult the Moodle site for the deadlines.
Send the files naming them as "Surname_Name_document.pdf" and "Surname_Name_slides.pdf/pptx".

- **Some data repositories** in Internet:
  - Bayesian networks: http://www.cs.huji.ac.il/~galel/Repository/
  - Probabilistic graphical models software:
    http://www.cs.ubc.ca/~murphyk/Software/bnsoft.html
    https://www.bayesfusion.com
    http://www.bnlearn.com/bnrepository/
  - KDNuggets competition site: www.kdnuggets.com/datasets/
  - UCI Machine Learning Repository: http://archive.ics.uci.edu/ml/
  - Kaggle competition data: http://www.kaggle.com/
  - At WEKA website: http://www.cs.waikato.ac.nz/ml/weka/datasets.html