

# Data Preparation in Machine Learning

## Scope and topics

Data preparation (DP) focuses on getting data with higher quality enabling a more accurate and reliable data mining model. There are many possibilities to cope with DP such as data selection to drop elements inside a dimension (such as the features or the instances), data imputation to recover missing values in classification, regression data or even in time series, data cleansing to detect values which differ outstandingly from the remaining data, etc. The computation time to get the final model after pre-processing may be reduced in the first case, increased in the second scenario and almost keep unaltered in the third situation, although the important thing is to compute the data mining model in a right direction instead of obtaining an initial model to be undergone to any kind of modification or even to be re-run many times given that the data are not enough ready to achieve a global and satisfactory prediction supported by the data. Tools based on graphical user interfaces are of particular interest in the sense that may make easier the procedure to prepare the raw data and eventually to get the ready data to face the mining phase. Submissions telling count on very recent and, if possible, unprecedented applications are encouraged. Additionally, new theoretical or empirical approaches are welcome. The topics of interest for this session include, but are not limited to:

- Data cleansing
- Data engineering
- Data fusion
- Data imputation
- Data pre-processing
- Data selection
- Feature selection
- Instance selection
- Outlier detection
- Data mining
- Motif mining
- Stream mining
- Itemset mining
- Sequential pattern mining
- Frequent pattern mining
- Infrequent pattern mining
- Rare pattern mining
- Signal mining
- Speech mining
- Text mining

## Organizers

*Antonio J. Tallón-Ballesteros (University of Huelva, Spain)*

*Luís Caviue (Universidade Aberta, Portugal)*

*Ireneusz Czarnowski (Gdynia Maritime University, Poland)*

## Contact information

*Antonio J. Tallón-Ballesteros*  
Higher School of Engineering  
Campus "El Carmen"  
Fuerzas Armadas AV.  
21007-Huelva (Spain)  
[antonio.tallon@diesia.uhu.es](mailto:antonio.tallon@diesia.uhu.es)