

Data Mining & Machine Learning Fundamentals



Mirco Nanni, ISTI-CNR

Ph.D program in Data Science @ Pisa
December 2019-January 2020



Calendar

- December 5, 9:00-12:00 (3h)
- December 11, 9:00-12:00 (3h)
- December 13, 9:00-11:00 (2h)
- December 19, 9:00-12:00 (2h)
- January 9, 14:00-16:00 (2h)
- January 10, 14:00-17:00 (3h)
- January 16, 14:00-16:00 (2h)
- January 17, 14:00-17:00 (3h)



Topics

Theory

- Data understanding and preprocessing
- Classification
 - Decision trees, KNN, elements of SVM & neural networks
- Clustering
 - K-mean, Hierarchical methods, DBSCAN
- Frequent patterns
 - Frequent itemsets, elements of Sequential Patterns & time series Motifs

Practice

- Hands-on exercises using Python (*)
- Standard libraries for data mining & machine learning

(*) Python 3 on Jupyter notebooks



Exam

To whom it may concern...

- Small DM/ML python project
- Dataset & objectives will be assigned during the course
- Spontaneous individual proposals for alternative projects are welcome
 - Requirement: apply (at least) an analysis method from each family



Mailing list

- Send an email to me to be enlisted:
 - mirco.nanni@isti.cnr.it
- Spam-free list, will be used for sharing material and last-minute communications