



# Course Introduction

Data Mining and Text Mining (UIC 583 @ Politecnico di Milano)

- ❑ Prof. Pier Luca Lanzi  
Dipartimento di Elettronica e Informazione  
[pierluca.lanzi@polimi.it](mailto:pierluca.lanzi@polimi.it)  
tel. 02 23993472  
<http://webpace.elet.polimi.it/lanzi>
- ❑ Office Hours  
Wednesday, from 15:00 until 17:00
- ❑ Teaching Assistant  
Dr. Daniele Loiacono  
Dipartimento di Elettronica e Informazione  
[loiacono@elet.polimi.it](mailto:loiacono@elet.polimi.it)

- ❑ Basic Introduction (24 hours)

  - Short introduction to Data Mining and Text Mining

- ❑ Advanced Techniques and Applications (16 hours)

  - Advanced Data Mining techniques and applications

- ❑ Final Project (24 hours)

  - An application to real-world data

- ❑ Shorter course covered by the first 20 hours
- ❑ Lectures are in English
- ❑ Final test can be either in Italian or in English upon request

- ❑ Introduction
  - ▶ Representation of data
  - ▶ The knowledge discovery process
  - ▶ Data mining
- ❑ Typical Data Mining tasks
  - ▶ Associations
  - ▶ Clustering
  - ▶ Classification
- ❑ Aggregate methods
- ❑ Preprocessing
- ❑ Advanced techniques and applications
  - ▶ Text Mining
  - ▶ Graph Mining
  - ▶ Data Streams
  - ▶ ...

- ❑ Students enrolled to the master at the Politecnico
  - ▶ Written test, 5 problems, 33 points total
  
- ❑ Students enrolled to the UIC-POLIMI master
  - ▶ Written test, 4 problems, 28 points
  - ▶ Course project, 5 points
  - ▶ 33 points total
  - ▶ The project involves data from an industrial problem and a set of questions that the students should answer using some of the techniques discussed during lectures
  
- ❑ 33 points corresponds to the grade “30 e lode”
  
- ❑ Students not enrolled to the UIC master who wish to do the course project should send an email to [pierluca.lanzi@polimi.it](mailto:pierluca.lanzi@polimi.it)
  
- ❑ This year the project will consist in the characterization of time series data

- ❑ The course page can be accessed at <http://www.dei.polimi.it/people/lanzi>
  
- ❑ Bibliography
  - ▶ Jiawei Han, Micheline Kamber. "Data Mining: Concepts and Techniques" Second Edition. Morgan Kauffman, 2006.
  - ▶ Ian H. Witten, Eibe Frank. "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations" 2nd Edition.
  - ▶ Tom Mitchell. "Machine Learning", McGraw Hill 1997
  
- ❑ Software
  - ▶ RapidMiner (Community Edition), <http://rapid-i.com/>