



FEP

FACULDADE  
DE ECONOMIA  
UNIVERSIDADE  
DO PORTO

# Data Mining I

**Bruno Veloso**

FEP & INESC TEC

2025/2026

# Instructors and Contacts

- **Coordinator:** Bruno Veloso (bveloso@fep.up.pt)
- **Contacts:**
  - LIAAD-INESC TEC, Campus da FEUP
  - Rua Dr. Roberto Frias, 378, Porto

# Syllabus (6 ECTS)

- Data Mining Tasks
- Basic Algorithms for Data Mining
- Classification
- Regression
- Evaluation of Data Mining Algorithms
- Pre-processing
- Multiple-Models
- Association Rules
- Methodologies for Data Mining

# Goals, Competencies and Learning Outcomes

At the end of the semester, students must:

- have knowledge about the different learning tasks;
- have knowledge about the main methods for each of the tasks;
- formulate a decision problem as a data mining task;
- be able to apply the methods to new data mining problems;
- be able to evaluate algorithms, interpret results.

# Evaluation

- **Home Work:** Kaggle competition
- **Final Exam**
  - 19 January 2026 18:30 and 30 January 2026 18:30
- **Final Grade:**  $\frac{\text{Home Work} + \text{Exam}}{2}$

# Evaluation: Kaggle Competition

- The work must be performed by groups of 5 students (individual works are not allowed).
- **Exploratory phase:** Exploratory data analysis of data: identifying outliers, anomalous examples, etc.
- **Predictive phase:** You can use any tool or combination of tools for the predictions (Python, Knime, R, Excel, Weka, or RapidMiner).
- Each group should submit a report until December 5th 2025 via Moodle.
- The report must have, at most, 12 pages.
- **Criteria:**
  - Rank of groups every week after 30 October.
  - Critical analysis of the results!
  - Argumentation and justification of the choices made.

# Resources

- Slides, papers, texts, and other material available in moodle:  
<http://moodle.up.pt>
- **Books:**
  - Library of Fep
  - Library of LIAAD
- **Video Lectures:** <http://videolectures.net>
- **KDD nuggets:** <http://www.kdnuggets.com>
- **Data Sets at UCI:** <http://archive.ics.uci.edu/ml/>
- **Data Sets at Kaggle:** <http://www.kaggle.com>

# Bibliography

- J.Gama, A.Carvalho, K.Faceli, C.Lorena, *Extração de Conhecimento de Dados*, Silabo, 3rd edition 2017
- Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*, 2nd edition, Morgan Kaufmann, 2006



# Bibliography - Complementary

- Ian Witten, Eibe Frank; *Data Mining: practical machine learning tools and Techniques with java implementations*, Morgan Kaufmann, 2000
- Tom M. Mitchell, *Machine learning*, McGraw Hill, 1997
- David Hand, Heikki Mannila, Padhraic Smyth; *Principles of Data Mining*, MIT Press, 2001

# Software

- **Knime:** <http://www.knime.org/>
- **Python:** <https://www.python.org/>

