



Universidade do Minho

**Aprendizagem Automática em Sistemas Empresariais  
(Predictive Business Analytics)**  
Mestrado em Engenharia e Gestão de Sistemas de Informação,  
Mestrado em Sistemas de Informação  
**Semester 1 2024/2025**

**PROJECT 1**

**Problem statement**

Using the Crisp-DM methodology it is intended to generate / induce one or several models / patterns for predicting, in the most efficient way, future values for a variable (target / label). Three approaches should be considered:

Supervised learning

i) Models for classification

ii) Models for regression

Unsupervised learning

iii) Models for clustering

One or more of the resulting models should be integrated into the analytical component of a business intelligence system to be developed in the context of project 2 (deployment phase).

**Data Set**

A data set is considered for competition:

<https://www.kaggle.com/datasets/shashanks1202/apartment-rent-data>

**Deadline for submission of final reports**

Week 14/10/2024 - P1 - Business Understand

Week 11/11/2024 – P1 - Data Understand & Data Preparation

Week 9/12/2024 - P1 - Modeling & Evaluation

**Working Groups**

4 elements (max.).

**Assessment Methodology**

To determine the grading, the following components and respective weights will be considered:

Report	20%;	(This parameter will be evaluated by applying a ranking among the working groups)
Application of the methodology	20%;	
Results/model performance	30%	
Continuous assessment	25%;	
Peer assessment	5%	

**Tools**

The elements of the working group will be consulted for a differentiation in the grades.  
The CRISP-DM methodology should be followed.

For data mining:

R (data mining package)  
Python  
WEKA  
RapidMiner  
Other

For data processing

MS-Excel  
DBMS (MS-SQL Server, MS-Access, MySQL, other)  
Other

## References

1. Santos, M.F., Azevedo, C. Data Mining - Descoberta de Conhecimento em Bases de Dados, FCA, Portugal, 2005.
2. Azevedo, Ana; Santos, Manuel (2008). KDD, SEMMA and CRISP-DM: a parallel overview; Proceedings of DM 2008 – IADIS European Conference on Data Mining 2008, pp 182-185.
3. Hastie, T., Tibshirani, Friedman J., The Elements of Statistical Learning – Data Mining, Inference and Prediction, Springer.
4. Han, J., Kamber, M., Data Mining: Concepts and Techniques, Morgan Kaufmann, New York, USA.
5. Berthold, M., Hand, D., Intelligent Data Analysis – An Introduction, Springer.
6. Crisp-DM, <http://www.the-modeling-agency.com/crisp-dm.pdf>
7. ML Mastery <https://machinelearningmastery.com>
8. DM tool Documentation.
9. Azevedo, Ana; Santos, Manuel (2009). An architecture for an effective usage of data mining in business intelligence systems. In Proceedings of the 13th IBIMA Conference on Knowledge Management and Innovation in Advancing Economies, pp. 1319 – 1325.
10. Azevedo, Ana; Santos, Manuel (2009). BUSINESS INTELLIGENCE - State of the Art, Trends, and Open Issues. In Proceedings of KMIS 2009 – 1st International Conference on Knowledge Management and Information Sharing, pp. 296-300.
11. Azevedo, A., & Santos, M. F. (2013). A Perspective on Data Mining Integration with Business Intelligence. In Data Mining: Concepts, Methodologies, Tools, and Applications (pp. 1873-1892). Hershey, PA: Information Science Reference. doi:10.4018/978-1-4666-2455-9.ch097
12. Business Intelligence and Analytics: Systems for Decision Support, Global Edition (10e), By Efraim Turban, Ramesh Sharda, Dursun Delen, Pearson Higher Ed USA, ISBN: 9781292009209
13. Business Intelligence, Analytics, and Data Science A Managerial Perspective, Ramesh Sharda; Dursun Delen; Efraim Turban, Pearson, ISBN: 9780134633282, 0134633288, Edition: 4th, 2017
14. Ana Azevedo, Manuel Filipe Santos (2015-2020). Integration of Data Mining in Business Intelligence Systems, IGI.
15. Machine Learning Algorithms: Popular algorithms for data science and machine learning, 2nd Edition, Giuseppe Bonaccorso, Pack Publishing, 2018.

## PROJECT 2

### Problem statement

It is intended to complete the deployment phase by building a Business Intelligence (BI) system based on data and data mining models resulting from project 1, allowing to achieve the goals defined in the phase of business understanding. The BI system should include:

- A database / Data Warehouse / Data Marts to store the data;
- An ETL component to populate the database (it can be in offline mode);
- An analytical component based on Data Mining models for predicting and a set of dashboards for visualizing management information and performance indicators (KPI).

### Important dates

Groups will be asked to make a short demonstration between 6 January and 10 January 2025. The report can be delivered by January from 13 to 19, 2025.

### Working Groups

4 elements (max.).

### Assesement Methodology

To determine the evaluation, the following components and respective weights will be considered:

Component	Weight
<b>BI System</b>	<b>70</b>
System architecture	10
Multidimensional model	10
Analytical component (data mining models)	20
Indicators	10
Dashboards	10
Quality and overall efficiency	10
<b>Demonstration</b>	<b>10</b>
Results presentation	5
Individual performance	5
<b>Report</b>	<b>15</b>
Peer assessment	5

The elements of the working group will be consulted for a differentiation in the grades.

### Tools

Groups are free to select from (not limited to):

- Suites - MS AZUR, MS SQL Server, Oracle BI
- Data Base Systems – MS SQL, Oracle, MySQL, MS Excel, MS Access, Oracle
- Visualization – Tableau, PowerBI

### References

16. BI tool documentation;
17. Scientific and technological documentation related to the application area.