

Programming for Data Science

Project

Objectives

- To strengthen students' ability to design and implement data-driven analyses using Python.
- To apply the CRISP-DM/POST-DM data science methodology systematically, from data collection to model evaluation.
- To develop proficiency in data preprocessing, exploratory analysis, and model interpretation using machine learning and network science techniques.
- To encourage critical thinking and methodological rigor in the analysis of real-world data.
- To enhance students' capacity for clear scientific communication, through well-structured code, reports, and visualizations.
- To foster originality and autonomy in selecting, justifying, and analyzing relevant data sources.

Project Description

Students are required to collect data from a relevant and credible source (e.g., <https://www.e-redes.pt>) or other open data repositories) and perform a comprehensive analysis applying machine learning and network science concepts. The project should demonstrate a full data science workflow: data acquisition, preprocessing, exploration, modeling, evaluation, and interpretation of results.

Deliverables

Each student or group must submit the following materials:

- Code files (`.py` or `.ipynb`) containing the complete implementation.
- Report (`.docx`) describing the methodology, analysis, and findings.
- Data files used in the project.

Report Structure

The report must include:

1. Abstract – Summary of objectives, methodology, and main results.
2. Introduction – Context, motivation, and project objectives.

3. Theoretical Background – Overview of key concepts (e.g., machine learning, network science, CRISP-DM/POST-DM).
4. Methodology – Description of data collection, preprocessing, modeling, and evaluation techniques.
5. Results – Presentation and analysis of findings, including visualizations and performance metrics.
6. Conclusion – Summary of contributions, limitations, and future directions.
7. References – Complete list of cited sources following an academic citation style.

Evaluation Criteria

Projects will be assessed according to the following dimensions:

- Programming and Technical Implementation (25%) – Correctness, functionality, and organization of code.
- Methodological Rigor (20%) – Consistent application of CRISP-DM/POST-DM and justification of analytical choices.
- Analytical Quality (20%) – Appropriateness of models, validity of results, and soundness of interpretations.
- Communication and Presentation (20%) – Clarity, structure, and visual quality of the report and results.
- Originality and Relevance (15%) – Creativity in data selection, problem framing, and analytical approach.