



## PROJECT AND RESEARCH METHODOLOGIES

### Master in Biomedical Data Science

After making a working team, each team must **follow the different Phases of Research methodology** with the given particularities explained below, and must do the following:

1. **Select the domain of research**,  
i.e. a Biomedical problem/topic and its corresponding bind dataset/s

**Once the domain has been selected (and some dataset/s have been selected) the topic must be communicated to the professor to get his approval.** Follow this steps:  
Prepare a one page report with the following information:

  1. Name of all group components by alphabetical order (sort by Family name)
  2. Topic
  3. 4 lines justifying the interest of the selected topic
  4. List of datasets to be involved and access details (url, etc)
  5. Send this document to the lecturers (Xavier Angerri and Karina Gibert) before the session
2. **Formulate the Research Problem / Initial Research Question(s)**
  - a) Define the Initial/tentative Research Problem
  - b) **Use the Canvas to design the scope and approach of your Research**  
(find the CANVAS template in the course web)  
**The CANVAS helps you to give form to different aspects relevant for the research**
  - c) Enumerate the initial goals/objectives of the Research problem
  - d) Make a risk analysis
  - e) Describe some KPIs
3. **Plan your research**
  - a) Rephrase the Research Problem into a working Research Question/Proposition(s)
  - b) Identify related keywords for literature search
  - c) Describe a Timeline for the Research Work (Gantt diagram)
  - d) Analyze consistency with the CANVAS and several fields in the CANVAS from a global perspective. Check the consistency of your research plan.



4. **Conduct the Literature Review/Background Research.** Specifically:
  - a) The group must **search for, at least, one research article for each member** related to the selected Biomedical problem/topic and/or dataset
  - b) Each member must **read and review her/his corresponding paper/s**
  - c) Each member must **analyse and find out/discover the research methodologies, research methods used and results obtained** in her/his paper/s
  - d) The group must **elaborate a state of the art of the topic (surveying the research methodologies used, research methods used, results obtained, drawbacks, advantages, etc.)** related to all the articles read.

5. **Refine the Research Problem Definition / Final Research Question(s) / Hypothesis Formulation**

- a) Define the Final Research Problem
  - b) Consolidate the final description of the goals/objectives of the Research problem
  - c) Clearly Formulate the Research Hypothesis/Hypotheses

6. **Identify the Relevant Information**

**Formulate the Research Design Document**

**Formulate the Research Project Design document:**

- a) Identify Relevant Information

- i. Data Selection

1. Identified datasets or other data gathering research methods
    2. Build metadata files
    3. Which data will be used
      - Basic structure of every data matrix involved:

Data source (url or company, etc)	Nr of records	Nr of variables	Nr of numerical variables (minimum 7)	Number of binary variables (minimum 2)	Number of other qualitative variables (minimum 5)	% of missing data in the dataset

- For each dataset involved in the previous table include: One paragraph explaining what data are about

Also, fill in the following template for each dataset involved in your system. It is possible to enlarge



Intelligent Data Science and  
Artificial Intelligence Research Center

your data base with additional variables coming from other sources if you like, but do not invest too much time on that

- Data Collection Design:  
Sampling/Experimental settings design

Data Source (url or company, etc)	Name of Variables	Type (Num, Qual, et)	Granularity/frequency (t real, online, ...)	Nature (Open, private, ....)	% of missing data of the variable	How to obtain the data in training step (basic download, on-line, query SQL, etc)	How to obtain the data in production step (basic download, on-line, query SQL, etc)

- Global idea about the knowledge components you think will be used in the system and information sources you intend to use with this components (literature, experts, etc)
- ii. Model Selection
1. Select relevant Data-driven models
  2. Select relevant Model-driven techniques
- b) Outline a Strategy to achieve the Research goals
- i. Selected Research Methodology Types
  - ii. An Initial Scheme/Chart/Architecture of the Research Project
- c) Initial Research Project Task Design
- i. Task Analysis
  - ii. Methods implementing each task
- d) Detailed Research Project Management Plan
- i. Project Management Methodology and tools
  - ii. Project Scheduling (Gantt diagram)
  - iii. Task Assignment
  - iv. Risk Management Analysis
7. Design and implement the data preprocessing process to prepare the data for the training: provide a global flow diagram for the entire *data preprocessing process*

-----SUPERVISION 2-----



8. Testing of Research Design on collected Data / Hypothesis Testing
  - a) Model Application
    - i. Induction of data-driven models
    - ii. If needed, use of model-driven techniques
    - iii. Model validation
  - b) Model Comparison / Hypothesis Testing
    - i. Null / alternative hypothesis
    - ii. Statistical tests

-----SUPERVISION 3-----

9. Analysis of results and Synthesis of conclusions.
  - a) Discussion of results
  - b) Conclusions
  - c) Future work
10. **Write the Research Proposal Document, a document summarising the proposed research, which must contain:**
  - a) Introduction/Proposed topic of Research
  - b) Literature Survey / State of the art
  - c) Motivation for the Research / Gap in the existing Research
  - d) Research Statement/Hypothesis or hypotheses
  - e) Specify the goals of the proposed Research Work/Project
  - f) Selected Research Methodology/ies to be used
  - g) Expected outcomes of the Research proposal
  - h) Plan of Research Work/Project (Gantt diagram)
  - i) References/Bibliography/Resources

At delivery date, you should deliver a **ZIP file** labelled as “**PRM-YearCourse-T#NumberOfTeamwork-TitleOfWork**”, should be uploaded through “**Campus Virtual**” in due time with the content bellow. It should include the corresponding parts and growth the materials along the course.

1. A folder named “**Documentation**” with:
  - a. The report
  - b. Gantt diagram with tasks planning of the ressearch development
  - c. Tasks assignment and responsibilities among teamwork members
  - d. Time sheet, with hours devoted by each teamwork member to each task, andglobal accounting by each teamwork member
2. A folder named “**Data**” with the files with the original raw dataset/s or database/s and preprocessed ones used for *data-driven models*, both for training and for testing.
3. A folder name “**References**” with the papers used for your research.
4. A folder named “**Models**” with the *data-driven models* or/and *model-driven techniques* used which can be visualized, read or interpreted in a nice way.
5. A folder named “**Source**” containing the source code of the scripts/macros/



## Intelligent Data Science and Artificial Intelligence Research Center

functions used in the different software tools/languages used at training step.

6. A folder named "**Presentation**" containing the slides (up to 15) of the public presentation of the work:
  - a. One title slide with the complete title of your work, your group number and names of all members of the group and date.
  - b. Several slides summarizing the report delivered. Be sure to include a slide with the kind of actors interacting with your research project.
  - c. Be sure to include a slide with the Canvas of your research.
  - d. Be sure to include a slide with the kind of actors interacting in your research.
  - e. Be sure to include a slide describing the 3 use cases tested.
  - f. Details on the different components of the system
7. A **README.txt** file specifying the structure and contents of the ZIP file.

Several items will be considered for the qualification of this work:

- The quality of the methodology and work done.
- The documentation delivered,
- The quality of the oral exposition (both presentation and content assessed, as well as the ability to answer questions),
- The planning, coordination and management of the team, and
- The individual evaluation of each student, including her/his integration level within the team group.