Instructions for the exam project

As a final fulfillment of the course requirements you will do the project on a data set of your choice. Here are stages of preparation and tasks you need to perform in order to be fully credited for the exam:

- Organize yourself into groups of two per group (there can be one group of three or one group of one person in the case of the odd number of students taking the class). The groups will be divided into three batches for the final presentation time (see the schedule): Groups 1-4 Exam 1, Groups 5-8 Exam 2, Groups 9-13 Exam 3.

- Find a data set and define a goal of the data analysis the would constitute the main theme of your project (there can be more than one data set but then they need to be related to one common general goal of the project). The following page for data repositories can be very useful: **top-sources-for-machine-learning-datasets**

- Prepare two pages description of what you plan to do and submit over the canvas for approval. The description should address the following:

  - Formulation of the main question that the project aims to answer.
  - Proposition of approaches to the problem and their rationales – based on the lectures and literature.
  - Some discussion of how the data set should be prepared or modified to meet the requirements of the methods, if applicable.
  - Approximate distribution of the efforts and tasks between members of the team.

- After the approval of the project theme and some possible modifications of it by the instructor, do analysis of the data.

- Organize the outcome of your analyses and in the form of slide presentation to be. In this presentation follow general instructions:

  - Presentation time is between 20min time plus one-two questions and a short discussion.
  - In the presentation, practical aspects of the project should be included including initial presentation of the data set with some visualization of them.
  - Formulation of the main question that the project aims to answer.
  - Brief account of methods used for the project and their rationales – based on the lectures and literature.
  - If encountered, any difficulties or problems associated with the data set should be reported including failed analyses.
  - Presentation how the actual analysis has been performed, including algorithms and utilized programs and/or packages.
  - Presentation of the results and their interpretation – graphs highly recommended if applicable.
  - Final conclusions from the study – both failures and successes. If failures it should be discussed what could be blamed for them and recommendations for future that would help to avoid the encountered problems.

If appropriate and desired you may go beyond the methods presented in the course. The presenters should be ready for questions and be well familiar with the methodology that has been used in the reported analyses. In all cases, the data should be at the center of the presentation and the conclusions should be more about the data than about the methods. Inconclusive results should be commented and some recommendation given.

Following are general rules about making presentation of your work on the final project.

Clarity of the presentation is the priority!

Finally, the presenters should count on that the audience (this includes the lecturer) is encouraged to exercise a 'scientific' directive as formulated in the quote

"There are naive questions, tedious questions, ill-phrased questions, questions put after inadequate self-criticism. But every question is a cry to understand the world. There is no such thing as a dumb question".

Carl Sagan,
(The Demon-Haunted World: Science as a Candle in the Dark)