AIDA
Project

Optimal data analysis pipeline

General information:

- A team of 3 students.
- Goals: 1) to conduct independent data exploration, and some research using machine learning methods, 2) to propose an optimal approach to process the data, and 3) to analyse and clearly explain the obtained results.
- Duration: 3 weeks
- Presentation of your results: you will be given 10 minutes to present your results (prepare slides!) + 10 minutes of questions

Project Schedule:

1. Milestone 1:
    - Download the data
    - Do some data pre-processing if necessary
    - Test some standard machine learning methods relevant for your problem
    - Read the papers related to the scientific problem (some papers will be provided but feel free to read more)
2. Milestone 2:
    - Analyse the results obtained with the state-of-the-art methods
    - Focus on an optimal approach. You can explore a combination of the existing methods, or a particular pre-processing or post-processing step, etc.
    - Implement it in Python. You are not asked to re-implement existing algorithms.
    - Explain why the chosen method is the best in your case (it can be the most accurate, or the simplest, or the fastest, or ...)
3. Milestone 3: Oral presentation. You are about 10 minutes to clearly present your results. I will ask you to send me your presentation (as a .pdf file). It is not needed to write a report. Your slides should include:
    - Problem description
    - Methods you tested and justifications why you choose them
    - The approach you consider to be an optimal solution for you problem
    - Implementation details, difficulties you met
    - Do not forget to mention relevant papers and published results
    - Conclusions and perspectives

Project Evaluation

- You should demonstrate that you well understand the methods you used and the results you obtained
- You have to prove it by comparing the performance of several machine learning methods
- Clarity of the presentation