Bonus Project for Data Science in Action

This document describes the bonus project for Data Science in Action. It is valid only for the academic year 2021/22. The project is optional and will give you up to 3 additional points towards your final grade in this class (Poor 0 points, Fair 1 point, Good 2 points, Excellent 3 points).

Instructions

Overview.
For your data science bonus project, you will design and investigate a novel idea involving all students in Data Science in Action. It will be a self-designed project, which you may carry out in consultation with the instructor (if you wish). The only requirements are that (i) you should train a model according to the data science pipeline we saw in class, and (ii) the topic of your project should be strictly related to this class, which implies that you should collect and use data about students in this class (i.e., you!). This is an opportunity for you to enhance your expertise on how to carry out a data science project (before doing a real project in the second semester).

Choose your team and register for the project.
You can choose your own team, provided that it consists of at most 3 people. To register for the bonus project, send email both to gitaliano@luiss.it and to mquerini@luiss.it with the names and emails of the team members by December 15, 2021. Late registrations will **not** be accepted.

When to submit your project.
You must submit your project by February 6, 2022, i.e., before the start of the second semester, i.e., by February 7, 2020. Late submissions will **not** be accepted.

What to submit for your project.
Each group must submit one PDF document describing the project and its working code (Python or R). It must be possible to easily run the code to test the solution delivered.
The PDF document should be 4-6 pages (excluding figures and references) and should include each of the following:

- Title and Team members
- [Section 1] Introduction – Describe briefly your project and motivate why you think it is important.
- [Section 2] Methods – Describe your proposed ideas (e.g., data collection, data cleaning, features, feature engineering, algorithm(s), training overview, design choices, etc.) and your environment so that:
    - A reader could understand why you made your design decisions, and the reasons behind any other choice related to the project
    - A reader should be able to recreate your environment
    - It may help to include a figure illustrating your ideas, e.g., a flowchart illustrating the steps you followed in your system(s)

- [Section 3] Code Description – Describe your code (at a high level) so that:
    - A reader could easily follow up your implementation
- [Section 4] Experimental Design – Describe any experiments you conducted in order to demonstrate/validate the target contribution(s) of your project; indicate the following for each experiment:
    - Main purpose: 1-2 sentence high-level explanation
    - Baseline(s): describe method(s) that you used to compare your work to
    - Evaluation Metrics(s): which ones did you use and why?
- [Section 5] Results – Describe the following:
    - Main finding(s): report your results and what you might conclude from your work
    - Include at least one placeholder figure and/or table for communicating your findings
    - All the figures containing results should be generated from the code
- [Section 6] Conclusions – List some concluding remarks. In particular:
    - Summarize in one paragraph what is the take-away point from your work
    - Include one paragraph to explain what questions may not be fully answered by your work as well as natural next steps for this direction of future work
- [Section 7] Bibliography (Optional) – Include any bibliographic reference needed

How to submit your project.

You are required to send your running code and your PDF report by email both to gitaliano@luiss.it and to mquerini@luiss.it by the given deadline.

Enjoy!