ISEG - Lisbon School of Economics and Management
Universidade de Lisboa

Course: Programming Techniques

Project

In this project, it is expected that the students collect data and use Python language to explore data, create a model and deliver the results. This work should be supported on Jupyter Notebook. A theme should be chosen, and data obtained using open data datasets (e.g. Pordata, European Commission, Lisbon Chamber, among others). Then one (or several techniques) should be applied. Examples of techniques that can be used: regression, logistic regression, random forest, cluster analysis, SNA, neural networks. Note that some of the topics may require more self-study than others.

Deliveries include:

- Report (report should be submitted in both formats: .docx filetype and PDF filetype)

- Jupyter Notebook

- Dataset(s) (including the source of those dataset(s))

- Eventually, students may also deliver results in the form of a web app developed in Flask

1. Report

The report should have the following structure:

I. Introduction

In the introduction, students must provide the context of the project. It is also important to identify the generic problem the students expect to solve and the main objective of the empirical work.

II. Literature review

The group should identify a small group of papers that may help to identify similar work developed by other authors.

III. Empirical Work

Empirical work should be based on a data science life cycle, as follows:

1. Data context

2. Data collection

3. Data preparation

4. Exploring Data

5. Data Modelling

6. Evaluation

7. Deployment

IV. Results and discussion

Results obtained from the empirical work should be compared to the results of other authors from literature review.

V. Conclusions

What was the purpose of the work? What do you conclude from literature and also from the empirical work?

References

All the project references must be in APA v.6 style. It is important that students use a reference management system, such as Zotero [https://www.zotero.org/]

Pages between 5 and 12 pages. The paper format must be followed.

2. Jupyter Notebook

All the data analysis steps must be clearly explained.

The team is encouraged to explore other techniques different from those presented in class. However, in order to be valued, the members of the group must master the techniques.

3. Datasets

Obtaining original, interesting and useful data is hard work. So, it is a task that is valued in the project. Data sources must be clearly identified. The process of data collection should be well described.

The easiest way of obtaining datasets is selecting datasets already used in published papers (sometimes published with the papers). Obviously, it should be avoided if you want to have a good mark. But it is possible.

4. Web App Using Flask

This is not a mandatory task.

5. Presentation:

14th November - Small pitch of the project to the Class (max 5 minutes). No necessary (although advised) all the members of the teams present the pitch.

4th/5th December – class final presentation. All the members of the team must present.