|  |  |
| --- | --- |
|  | **ISEG - Lisbon School of Economics and Management** **Universidade de Lisboa****Course: Programming Techniques** |

**Project**

In this project, the students must collect data, use Python language to explore data, create a model and deliver the results. Jupyter Notebook supports the project. Students must choose a subject, and obtain data using open data datasets (e.g. Pordata, European Commission, Lisbon Municipal Government Site, among others). Students must apply one (or several) techniques. Students may use the following techniques: 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 (students must submit the report 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

Pay attention to plagiarism. ISEG Rules will be strictly followed.

**1. Report**

The report should have the following structure:

1. **Introduction**

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

1. **Literature review**

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

1. **Empirical Work**

To make the empirical work, students must follow a data science life cycle (like CRISP-DM or POST-DS), as follows:

1. Data context

2. Data collection

3. Data preparation

4. Exploring Data

5. Data Modelling

6. Evaluation

7. Deployment

1. **Results and discussion**

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

1. **Conclusions**

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

**References**

All the project references must be in APA v.6 style. It is essential 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**

The report must explain clearly the data analysis steps.

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

**3. Datasets**

Obtaining original, curious and useful data is hard work. So, it is a task that is valued in the project. Students must identify the data sources. 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 articles). Obviously, students must avoid if they want to have a good grade. But it is possible.

**4. Web App Using Flask**

Web app using flask is not a mandatory task.

**5. Presentation:**

19th November: Small pitch of the project to the Class (max 5 minutes). It is not needed (although advised) that all the members of the teams present the pitch.

16th/17th December: Final presentation. All the members of the team must present.