

Mini project 02 (Oral presentation)

Dataset

Each team must select an original dataset. The dataset can be directly downloaded from the Internet, or for a better score the team can generate the data based on different sources or directly generate the data.

Required sections (besides the usual sections: abstract, content, etc.):

- Perform a basic data analysis describing the dataset, summary statistics, data distribution, etc.
 - Describe the data domain. A complete and deep explanation.
 - How the data was recollected, limitations of the study, disadvantages, etc.
 - Describe the distribution of the data.
 - A few, but interesting plots.
- Preprocessing
 - Compute and describe all the steps needed to transform the data into a suitable dataset for the algorithms.
 - Steps needed to transform raw data into a suitable dataset.
 - Missing values.
 - Normalization.
 - Other processes.
- Processing and results
 - Clearly explain the algorithm
 - Compute and describe all the phases used in the algorithm (the team must also explain how kNN works)
 - Classification outputs.
 - Frequency tables and interpretation.
- Conclusions and limitations.
 - Does the study generalize to other domains?
 - Limitations.
 - Advantages.

Considerations:

- All the previous points are present in the report.
- Remember that the format of this project is an oral presentation (only the video, no report is needed).
- Slides must be in English.
- All the code must be visible.
- Quality of the presentation.
- Reproducibility.
- Clear description of each step. Ex. How the students managed missing values
- Originality.
- Similar presentations are eliminated (without further questions).
- The interpretation of the results is a key point to evaluate. Students should give clear and deep explanations of each point.
- Each graphic or table should be fully explained.
- Tables and graphics must be referenced and have their corresponding caption.
- Aesthetics of the presentation. An adequate size of graphics and tables. Nice merging of graphics, tables and text.
- First slide with the title, second slide with an abstract, third page an index, last slide the bibliography.
- The students, optionally, can upload their presentation to Rpubs (without the video).