Case-study guidelines

Aim of the case-study is to experiment with data and compare different classification methods:

- *collect the data*
- *clean/preprocess the data, define new features if needed*
- *experiment with different classifiers*
- *draw conclusions*

1. Work in teams of **three** students.
   Please, let me know your choice by 19.04.2021 (e-mail with one list of all the teams and their members).
2. Each team gives two presentations:
   - preliminary on **10 May 2021**
     - it should focus on a detailed description of the selected dataset and its characteristics; give examples from the dataset and the interpretation of attributes, show all you have learned about the dataset and the attributes (attribute types, meaning, histograms, correlations etc), explain the posed classification question, show performed data cleaning and preprocessing ...
   - final on **7 June 2021**
     - it should focus on a detailed description of all the applied classifiers (**at least three**) and their parameters, the used software/tools, show and interpret all the interesting knowledge induced by those approaches, compare the approaches…
3. Each presentation should be 8 minutes long. The presentations will be graded. Please send me the slides from the presentations right after they are given.
4. Each team should choose a dataset for a **classification task**.
   You can have a look at:

   https://www.kaggle.com/datasets

   https://archive.ics.uci.edu/ml

   https://pub.towardsai.net/best-datasets-for-machine-learning-data-science-computer-vision-nlp-ai-c9541058cf4f
5. The performed case-study should be your original work.
6. You can use all the tools, environments, (visualization) techniques etc. you like.