PROJECT: Introduction to Machine Learning

Rauzan Sumara

13/10/2022

1. Project should be carried out in a group (2 students)

2. Choose a dataset suitable for you based on the repository below:

   - "UCI machine learning": https://archive.ics.uci.edu/ml/datasets.php

   - "Mendeley Data": https://data.mendeley.com/research-data/?

3. The selected dataset should have **at least 1000 observations** and **at least 5 attributes/variables**.

4. Things to do in the project:

   a. Description of the dataset and your goal/objective

   b. Look at big picture of dataset, descriptive statistics and visualize the data (data exploratory analysis)

   c. Prepare the data for machine learning project

   d. Select and train the model (you can take into account one machine learning algorithm or more)

   e. Evaluate the model (i.e. check the following measures of classification: accuracy, sensitivity, specificity, F1, etc)

   f. Conclusion

5. The project can be delivered as the **Word file (.docx/.pdf)**, **Jupyter Notebook (.ipynb),** or **R Markdown (.Rmd).** In the project, you are allowed to use either Python or R programming.

6. Please pay an attention to the quality of your presentation. Your steps should be clearly explained, along with a description of the data used in the project, with the methods used, and the results of the analysis along with the conclusions.

7. The projects should be delivered after presentation via email: **rauzan.sumara.dokt@pw.edu.pl** with the email title:

   "IML Lab Project 2022 – [Your Group Number]"