DPRPy 2024/2025

Homework assignment no. 2A (max. = 15 p.)

Maximum grade: 15 p.

Homework should be sent via the `MS Teams` platform. You should send **1 file** containing solutions to tasks, i.e. `Last-name_First-name_HA_2A.ipynb` - a file prepared with `Jupyter Notebook`.

Remember to comment your code and take care of the overall readability of, both, your code and a file itself.

1 Tasks description

Consider *Student Performance* [Cortez, P. (2008). Student Performance [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5TG7T.] data set that contain detailed information regarding student achievement in secondary education of two Portuguese schools. The data attributes include student grades, demographic, social and school related features) and it was collected by using school reports and questionnaires. Two datasets are provided regarding the performance in two distinct subjects: Mathematics (mat) and Portuguese language.

See more at:

- https://archive.ics.uci.edu/dataset/320/student+performance
- https://www.semanticscholar.org/paper/61d468d5254730bbecf822c6b60d7d6595d9889c

Data can be downloaded from:

https://archive.ics.uci.edu/static/public/320/student+performance.zip

For Task 2 and 3 include in your solution short commentary about results you got.

1.1 Task 1 [3 p.]

Create `students` data frame in Pandas by creating one data frame that consist of merged information given in data set `student-mat` and `student-por`. Change the names of selected columns if needed.

1.2 Task 2 [6 p.]

1. Create a two new features which will convert final grades from math and portugal class to scale Polish grade system, i.e. $\{2, 3, 3.5, 4, 4.5, 5\}$ according to following conversion table:

% of maximum grade	Grade
$[0, 50]$	2,0
$(50, 60]$	3,0
$(60, 70]$	3,5
$(70, 80]$	4,0
$(80, 90]$	4,5
$(90, 100]$	5,0

2. See how the amount of free time after school affects students performance by calculating minimum, maximum, median and average of the final grade in groups determined by the amount of free time after school.
3. Calculate the correlation between number of past class failures and the number of school absences.

1.3 Task 3 [6 p.]

Try to answer following questions based on some appropriate data grouping / aggregation.

1. How health status of a student and quality of his/her family life influence the overall performance of the students?
2. Does having the access to internet affects the students grades and his/her desire to pursue higher education?