

Data Science 2

Homework 2021/2022 General Instructions

General

- Four assignments will be given. It is necessary to successfully solve all of them.
- Students are expected to collaborate in groups of up to two people. Only one solution per group is then handed.
- Assignments will be published in the repository <https://github.com/wenceslasdk/data-science-2/blob/main/Homework> and notifications via mail will be sent (please check that the mail contact in the University system is up to date)
- Students will have two weeks for successful submission. The dates of publication will be set according to the development of the tutorials (first will be given approximately in the fifth week of semester)
- Contact for any homework-related queries: Mgr. Ondřej Týbl, tybl@karlin.mff.cuni.cz

Successful submission

Typical assignment will ask the students to perform some prediction using specified method. For every assignment we expect two outcomes

- Submission of prediction to Kaggle competition. Link to the competition will be specified for each of the assignments together with instructions on the format. Only predictions that give a better evaluation metric than the given threshold will be considered.
- Readable python code with reasonable comments that leads to the desired prediction. This code is expected to be uploaded into module Moodle page in the specified due date in the following format: 2022_fullname1_fullname2_HW_X.py, where fullname1, fullname2 are full names of the students in the group (or 2022_fullname_HW_X.py if no collaborators) and X is the order of the assignment.

Successful completion

Each assignment may be returned to the students with some questions/remarks and the students must incorporate those into their solution within one week and send the correction, again uploaded to Moodle page (now using suffix _correction). The assignment is not successfully completed until no questions/remarks remain unsolved.