

Rules for completing coursework

One of the following assignments may be completed as coursework.

- 1) Solving a machine learning problem on a real dataset. The source of such datasets can be real data, data from Kaggle competitions and other similar platforms. 2) Implementation of the machine learning algorithm from the article, for which there is no reference or commonly used Python implementation. The presence of some implementation is not a ban on the implementation of the project, but in this case a comparison must be made between your implementation and the one given. 3) Some kind of your own project, if you discuss it with the teacher.

Work may be performed in groups of up to three people. In the case of more than one participant in the group, for each area of work there must be a person who is able to clearly explain what is happening in this area and vice versa, for each participant there must be a significant fragment for which he is responsible.

Deadlines

1. The defense of the work will take place at the end of December. 2. Deadline for choosing a project (dataset, article,...) and combining into groups is the end of November.

Filing an application

Write a letter to the teacher, in which you briefly describe what you plan to do (dataset, article, some ideas on methods) and the composition of the participants.

Requirements for work in the form of solving a machine learning problem

1. The dataset must be large enough (at least 10 thousand examples) 2. The number of tested methods for solving the problem must be at least $1.5N$, where N is the number of people performing the work. 3. For the task, a method for measuring the quality of the model must be selected and explained its applicability in this case. 4. Visual analysis of the data must be carried out. 5. For each method, data preprocessing must be performed or it is shown that it is not required. 6. Hyperparameters must be selected for each method. 7. For a regression problem, an analysis of the target variable must be performed and a transformation applied or an explanation of why it is not required. 8. The solution must be reproducible.

Requirements for work in the form of an implementation of the algorithm

1. The theoretical assumptions behind the algorithm must be analyzed
2. The correctness of the implementation must be shown (tests on simple cases)
3. A theoretical analysis of the computational complexity must be carried out algorithm and required memory.
4. There must be a comparison with other methods on the task and some kind of dataset.
5. The operating time and memory consumption should be measured depending on the size of the dataset (you can take subsamples of one fairly large dataset).