FACULTY OF COMPUTER AND INFORMATION SCIENCE

# Assignment #5

Course: *Machine learning*
Date: *November 25th, 2024*

### Assignment

In this assignment, you will learn about gradient boosting.

Write the code for gradient boosting of trees to solve a binary classification problem from scratch.

Use the Scikit-learn regression tree implementation. The trees have to be weak learners, so their depth has to be set to a small value. You can choose their final depth arbitrarily. Think about overfitting. Watch the video: `https://www.youtube.com/watch?v=jxuNLH5dXCs`

Download the dataset on Učilnica "wine-quality.csv".

This is the same dataset as used in Assignment 3, but now we are solving a binary classification problem (quality is the target and is either high or low). Be careful about continuous and categorical variables.

Test different tree depths. The trees have to be weak learners.

Test different learning rates. What is a good learning rate?

Test different numbers of trees that are built during gradient boosting and comment on the results. Does your model overfit? If yes, try to prevent overfitting.

You have to explain your solution.

Compare the results from your implementation with the "GradientBoostingClassifier" classifier implemented in Scikit-learn.

Be careful to use the same cross-validation or train/test splits for both implementations.

Try modeling your data also with XGBoost and try to find the best set of hyperparameters.

When you perform hyperparameter tuning be careful to use a validation set.