# 02. Regression – Homework

## Overview

* Regression analysis of Air Quality dataset:  
  [https://archive.ics.uci.edu/ml/datasets/Air+Quality#](https://archive.ics.uci.edu/ml/datasets/Air+Quality)
* Target - C6H6(GT)
* -200 indicates missing values

## Evaluation Criteria

Scoring (10 max):

* EDA: exploration of variables and properties of data with conclusions - 1
* Data preparation: all missing values are filled/dropped with explanation of why some method was chosen - 1
* Data preparation: normalization of data / scaling - 1
* Baseline model - linear regression without regularization - 1
* Feature engineering - basic transformations (nonlinear) - 2
* Metrics chosen as well as reasoning behind each metric - 1
* Feature importance, hyperparameters tuning - 1
* Statistics material used (residual analysis, factor-importance hypothesis testing) - 1
* Non-classic regression (SVM-regression, Kernel-regression, SGD-regression) – 1

Критерии оценки

Подсчет очков (не более 10):

• EDA: исследование переменных и свойств данных с выводами - 1

• Подготовка данных: все пропущенные значения заполняются / отбрасываются с объяснением того, почему был выбран какой-либо метод - 1

• Подготовка данных: нормализация данных / масштабирование - 1

• Базовая модель - линейная регрессия без регуляризации - 1

• Особенности проектирования - базовые преобразования (нелинейные) - 2

• Выбранные метрики, а также обоснование каждой метрики - 1

• Важность функции, настройка гиперпараметров - 1

• Использованный статистический материал (остаточный анализ, проверка гипотез важности фактора) - 1

• Неклассическая регрессия (SVM-регрессия, Kernel-регрессия, SGD-регрессия) - 1