

The task is based on an energy efficiency prediction problem which is a regression problem. More specifically, you are required to perform an analysis using different building shapes, with different characteristics, and predict the heating load of the building. The buildings differ with respect to the glazing area, the glazing area distribution, the orientation, and other aspects comprised in the dataset.

The dataset (below) for this task includes 9 features, denoted by X0, X1, ..., X8, and an outcome variable Y which needs to be predicted. The dataset contains missing values. Here is the meaning of the 10 variables:

X0: Category of the building
X1: Relative Compactness
X2: Surface Area
X3: Wall Area
X4: Roof Area
X5: Overall Height
X6: Orientation
X7: Glazing Area
X8: Glazing Area Distribution
Y: Heating Load

You are required to perform this analysis comprising: data inspection and visualisation, data preprocessing including data splitting in 70% training data, and 30% test data, data transformations you consider useful for this task, treatment of missing values, feature selection if you consider it useful for helping you achieve a better performance, etc. The analysis should include developing the predictive models based on the following algorithms already studied in this module, or that are going to be studied such as neural networks: simple Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, Polynomial Regression with regularisation, and Neural Network. These models (except simple Linear Regression which needs only to be simply trained) will be tuned using the training set. The training set will be used to select the best 2 models. Only these 2 best models will be evaluated on the test set. You can use any Python library.

The code, comments and explanations will be provided in a Python Jupyter notebook, which should include also the results of running your notebook.

(You can read the dataset simply using pandas function `read_csv` as follows: `data = pd.read_csv("datcw_na.csv")`)

Note regarding working in a team or individually, and what you need to submit:

- **You can work and submit in a team of 2 students** - in which case you should choose a team leader. As a team you should tackle the task above entirely.

Include the names and student numbers of both of the team members on top of the Jupyter notebook. The team leader must perform the submission from their account (hence only once per team, for both students) on VLE. (You shouldn't submit twice, in each student's account.)

- **Or you can also work and submit alone** for this assignment. In this case you are required to choose and develop only **3 models** out of the Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, Polynomial Regression with regularisation, and Neural Network. Include your name and student number on top of the Jupyter notebook followed by the mention "**I worked and submitted alone**".