**Task Title:** Housing PricesPrediction using Machine Learning

# Deliverables:

* A Python file with your complete implemented pipeline.
* A report summarizing your findings and discussing the performance of your model. This report should include the same sections as the [detailed requirements section](#bookmark=id.m78mnav35bzu) with the mention of the techniques you applied in each -code snippets are allowed.
  + The report **must contain** your learning curve(involves plotting two key metrics (such as error or accuracy) against the number of training examples or the training iteration. This visualization helps to understand how the model is learning and whether it is overfitting or underfitting.)

# Assignment Deadline:

Monday 17/03/2025 at 11:59 PM

Deliver your work in the “Task Submission “on moodle.

# Description:

In this assignment, you will be building a machine-learning model to predict house prices based on various features of the properties.

Your task is to build a regression model that can accurately predict the price given its features. You will start by preprocessing the data by handling missing values, scaling the features if necessary, **and splitting it into** **training , validation and test sets**. You can use k-fold technique.

You should split the data according to what you studied in the lectures into:

* Training set: to train the model
* Validation set: to evaluate the model and tune it
* Test set: to see performance on new data points (Generalization)

Next, you will implement and train a regression algorithm such as Linear Regression, Polynomial Regression, Ridge Regression or Lasso Regression. You will evaluate the performance of your model using metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared score on the test set.

Finally, you will select the best-performing model and fine-tune its hyperparameters using techniques such as Grid Search and Random Search**(self-study).**You will then evaluate the final model on the test set and report the final performance metrics.

# Detailed Requirements:

Follow this section thoroughly to satisfy all the needed requirements of this assignment.

## 1 Data Preparation

### 1.1 Missing Data

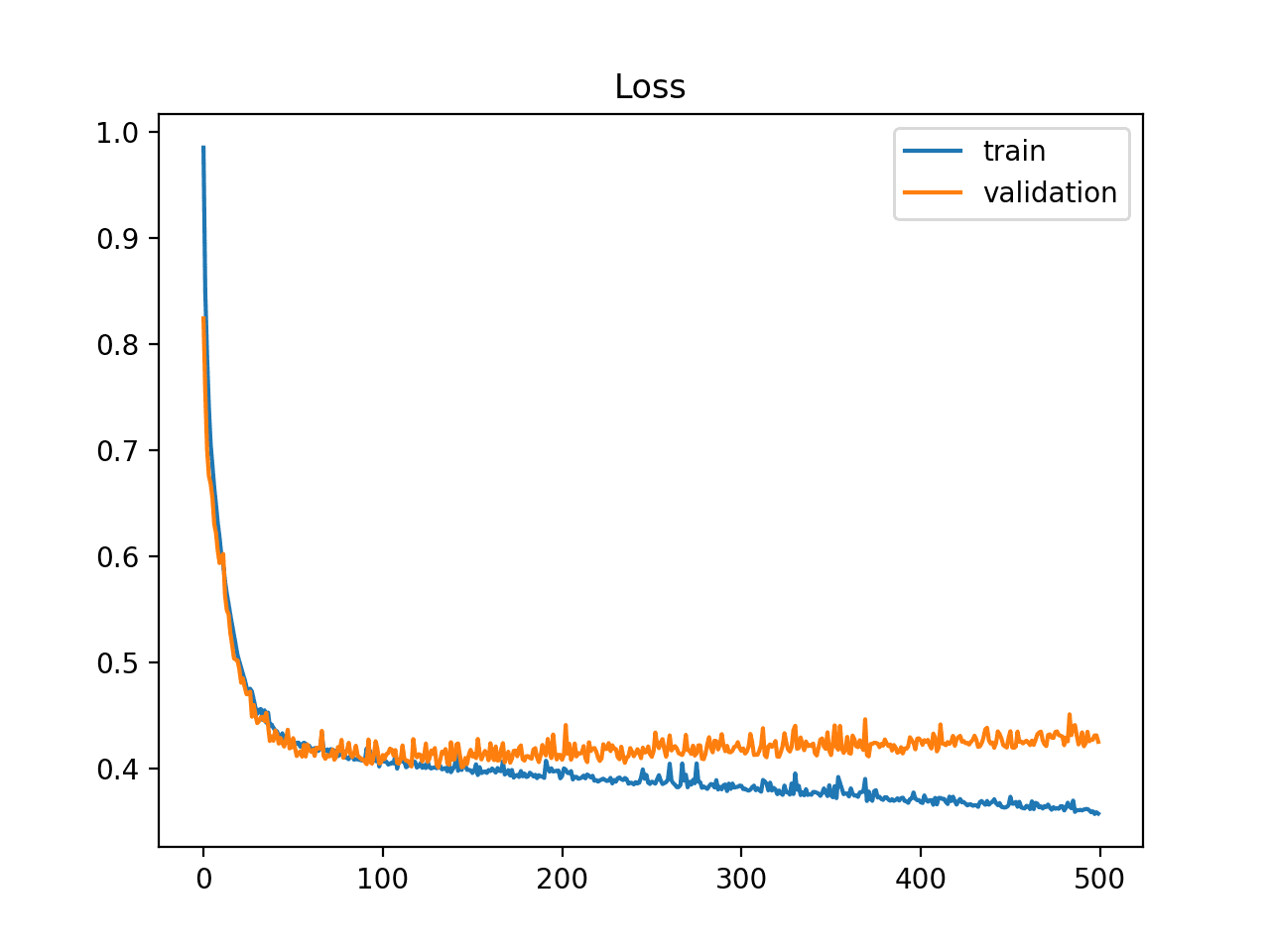
* Report your analysis of the missing data % for each attribute.
* Apply a data imputation technique suitable to each attribute with missing values.

### 1.2 Feature Engineering

* Encoding the categorical columns.
* Feature Scaling
* Feature Selection

## 2 Machine Learning Model

### 2.1 Plotting the Learning Curve



In this requirement, you are tasked to plot the Loss function (MSE) of both your training subset and validation subset over the epochs.

You can either implement a vectorized linear regression model and track your loss function per epoch, or you can utilize the Scikit-learn modules (<https://scikit-learn.org/stable/modules/learning_curve.html>).

## 2.2 Hyperparameter Tuning

Tune at least one of the model’s hyperparameters and report your findings.

## 3 Testing during discussion phase

After choosing the model and saving it, prepare a function that applies all “Data Preparation” process on any new data that we will provide for you during the task discussion.

During the task discussion we will provide you with a never- seen-before TEST data (In the same format as training data) and then make a prediction using your chosen model.