# Lecture 2: *"End-to-end Machine Learning Project"*

## Book Chapter

Stratified Shuffle Split

split the dataset into train and test sets in a way that preserves the same proportions of examples in each class as observed in the original dataset. This is called a stratified train-test split

Please read the *Chapter 2* **"End-to-end Machine Learning Project"** and answer the following Questions.

**Keep in mind: If you answer these questions and write a detailed summary, you won't need to read these chapters again while preparing for the exam**

## Questions

1. What are the 8 basic steps for an End-to-End Machine Learning Project?
2. What is a "pipeline"?
3. What is MAE and RMSE? Mean Absolute Error, Root Mean Squared Error
4. What can you see in the data using a histogram? Distribution of data
5. Why can a stratified shuffle split be useful?
6. What does the "standard correlation coefficient" state? linear relationship of 2 variables in range of -1 to 1
7. What does the scatter matrix on p. 60 reveal? Map of California
8. Why is data cleaning important? Garbage in garbage out
9. What does "one-hot encoding" mean and when do you use it? The conversion of categorical information into a format that may be fed into ML to improve prediction accuracy
10. What is the difference between normalization (min-max-scaling) and standardization? Standardization is the subtraction of the mean and then dividing by its standard deviation. Normalization is dividing of a vector by its length and transforming it into a range between 0 and 1.
11. What can be causes of an underfitting model? A simpler model for the chosen use case.
12. How does Cross Validation work? Resampling of the data by using different bits to test and train a model on different iterations
13. What does GridSearch do?

Randomized Search

## Homework Assignment

Have a look on the Jupyter Notebook for this chapter and do the tasks in the excercise section (on your own).

Please download the, unpack 02-End_To_End_ML_Project.zip it and run it on your machine.

Please accomplish the tasks mentioned there.