

# Assignment

## We'll cover the following

- Practice Makes Perfect: Working With Real Data
- Open Datasets

## Practice Makes Perfect: Working With Real Data #

Now you have a good idea of what a machine learning/data science project looks like. You have gained familiarity with great tools and techniques that you can use to train ML models. As it should be pretty obvious by now, much of the work is in the data preparation step. In fact, first-timers are often surprised by how little time in a machine learning project is spent actually doing machine learning. But it makes sense if you consider how time-consuming it is to gather data, integrate it, clean it, and pre-process it, and how much trial and error can go into feature design. Machine learning is not a one-shot process of building a dataset and running a learner, but rather an iterative process of running the learner, analyzing the results, modifying the data and/or the learner, and repeating.

The machine learning algorithms are important but, when getting started, it is recommended to be comfortable with the overall process first and learn just a few algorithms well, rather than spending all your time in learning advanced algorithms at the cost of ignoring the overall process.

### Your Turn Now!

1. First, try to improve the performance of the model for the housing dataset by using different models, selecting different features, replacing GridSearchCV with RandomizedSearchCV, trying out a different set of algorithms, etc.
2. Then select a dataset from a domain of your interest and go through the whole process from start to end. The key is to **practice, practice and**

then some more practice!

## Open Datasets #

There are thousands of open datasets, ranging across all sorts of domains, just waiting for you. Here are a few popular places you can look at to get lots of open data:

- [Kaggle datasets](#)
- [Amazon's AWS datasets](#)
- [Wikipedia's list of Machine Learning datasets](#)

I would recommend you start on [Kaggle](#) because you will have a good dataset to tackle, a clear goal, and people to share your experience with.

***Looking forward to hearing about all your great projects, and progress!***

