# Introduction

In this case study, you will be completing a real industry machine learning project. The project involves making sales predictions (sales forecasting) for various stores of a large retail corporation, based on a variety of different features. Along the way, you will also learn techniques for efficient data processing and model training.

A. Industry ML projects

Machine learning projects in industry are comprised of five main parts: data analysis, data processing, creating a model, training the model, and interpreting results. Data analysis is used to confirm that machine learning can actually find useful trends in the dataset to make the requisite predictions. Oftentimes, people will forget to perform data analysis and end up confused why their machine learning model doesn't work well, when in reality it is because the dataset is not conducive to making predictions.
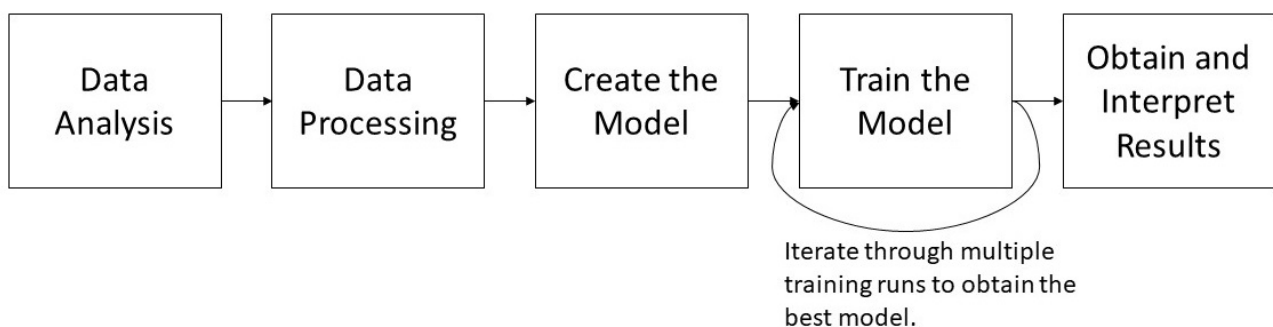
The timeline for an industry ML project.

After data analysis and data processing, we create the appropriate machine learning model given the application and dataset. In most cases, a regular multi-layer perceptron (MLP) will suffice, but sometimes we use a convolutional neural network (CNN) for images/videos or a recurrent neural network (RNN) for text-based data. Creating the model is the easiest part of an

industry ML project, since you just need to reproduce the code for one of these well-known models..

Finally, after training the model on the dataset, we use it on test data to confirm that the model works. The results obtained from using the model on the test data will be interpreted and reported to the project's supervisors. This is where we can confirm to the supervisors that the model indeed works for the given application, and can be deployed to production.

B. Sales forecasting

One of the most common uses of machine learning in industry is to predict retail sales for consumer-based corporations. Our case study will have you mimic the work of an ML engineer at a large retail corporation, where you'll be creating a model to predict the weekly sales of stores, based on a large training dataset.

In this section of the case study, you'll be performing data analysis to get a feel for the dataset and also confirm that there are enough trends in the dataset to predict weekly sales. You'll also perform a bit of data processing to organize the raw dataset, although the majority of the data processing is done in the **Data Pipeline Creation** section.