# A2 Research

## Semi-Supervised Learning

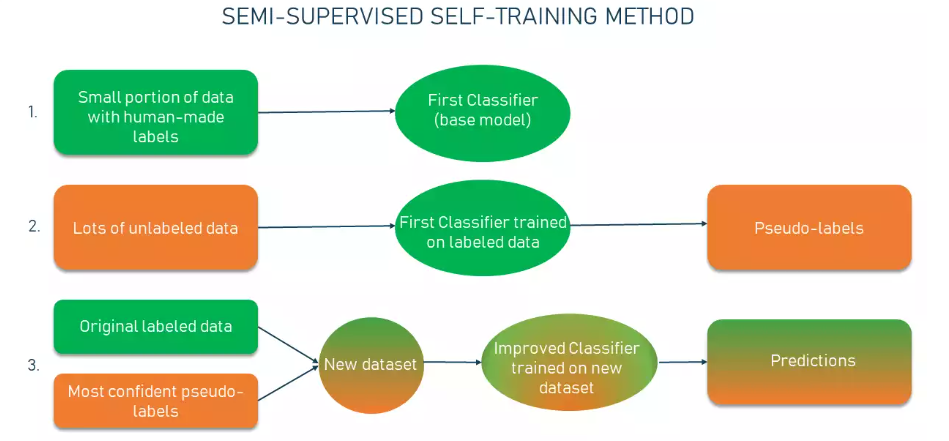
Also called Weak Learning

<https://www.altexsoft.com/blog/semi-supervised-learning/>

Technique to train using both labelled data and also unlabelled data.

The basic premise is this:

1. Train a base model on the labelled data
2. Apply **pseudo-labelling:** 
   1. This is where you predict on some or all of the unlabelled data.
   2. Then, you try to rate the confidence of your predictions.
   3. Find the most confident of the predictions, for example greater than 80% confidence
   4. Take these most confident predictions and combine them with the original labelled dataset.
   5. Then, train a second model, based on both the labelled and most confident predicted data
   6. This process can iterate (10 rounds is common)



There is another similar process called **Co-training.** Haven’t read yet into the details but apparently it’s about training two individual classifiers based on two views of the data