**Becoming a Machine Learning Engineer | Step 2: Pick a Process**

Picking your process is super important

After a few applied machine learning problems, you usually develop a pattern or process for quickly getting started and achieving good results. Once you have this process it is trivial to use it again and again on project after project. The more developed your process, the faster you can get to results!

经过几次应用机器学习的问题，您通常会开发一个快速入门的模式或流程，并取得良好的效果。一旦你有了这个过程之后，再重复地在项目中使用它是非常简单的。这个过程越来越熟练之后，你会更容易得到想要的结果！

Let me give you a head start and teach you a 5-step systematic process that I developed while becoming a machine learning engineer. This is just a starting point and you should feel free to change it to suit your needs

现在让我开个头，教你我在作为一个机器学习工程师的时候，使用的5步骤过程。这里仅仅是抛砖引玉，还需要你根据自己的需求修改它。

**Define the problem**

**定义问题**

This step is all about learning more about the problem at hand. Familiarize yourself with the domain and understand why you are building this solution. To help facilitate this, always ask yourself the questions below

这一步是要了解更多关于手头的问题。熟悉相关领域，并了解为什么要构建此解决方案。 为了帮助实现这一点，请总是问自己下面的问题：

What is the problem? Describe what the problem is formally and informally. Make sure you list assumptions you are making and any problems that are similar

**问题是什么**？正式地和非正式地描述问题。确保列出你正在做的假设和任何类似的问题。

Why does the problem need to be solved? List any motivations for solving the problem. What are the benefits a solution brings and how would you use it?

为什么要解决这个问题？列出解决问题的动机。解决方案带来的好处是什么？如何使用它？

How would I solve the problem? Describe how the problem would be solved manually to build up domain knowledge

我将如何解决这个问题？描述如何手动解决问题以建立相关领域知识。

**Prepare Data**

**准备数据**

Do you understand the data you have been given? Lots of people skip over this step because it is often tedious but it is super important. This work forces you to think about the data in the context of the problem before it gets lost in the craziness of algorithms

你了解你所得到的数据吗？很多人跳过这一步，因为它经常是乏味的，但它是非常重要的。 这项工作迫使您在数据迷失在算法的疯狂之前，能够在问题的背景下思考数据。

**Data Selection：**Consider what data is available to you. Is there any data missing? Can you remove any data?

数据选择：考虑可以使用哪些数据。有没有数据丢失？你可以删除任何数据**？**

**Data Preprocessing:** Organize your selected data. Format is, clean it, and take a sample from it

数据预处理：组织您选择的数据。格式化它，清理它，并从中取样。

**Data Transformation:** Processed your ready data for machine learning by engineering its features using scaling, attribute decomposition, and attribute aggregation.

数据转换：通过使用缩放，属性分解和属性聚合来获取其特征，从而处理为机器学习而准备好的数据。

**Explore different Algorithms**

**探索不同的算法**

Now that you have your data it’s time to try out a bunch of different standard machine learning algorithms. Typically, you would run 10–20 standard algorithms on the transformed and scaled versions of the dataset you prepared in the last step.

现在您已经掌握了数据，现在可以试用一些不同的标准机器学习算法。通常，您将在上一步中准备的转换后的数据集上运行10-20个标准算法。

The main goal of trying all of these different algorithms and dataset combinations it spreading your net far and wide. See what works and what doesn’t then go from there. More detailed explorations will follow with well performing algorithms.

尝试所有这些不同的算法和数据集组合的主要目标是将您的网络广泛传播。 看看有什么作品，什么不从那里去。随着更好的执行算法，将会有更详细的探索。

**Improve Results**

**改善结果**

After you have finished exploring the different algorithms and picked one that works well for your dataset it is time to squeeze out the best results from it. You can do this in a few ways, but it’s important to make sure that your results are significant at this point because hyper-parameter tuning isn’t going to turn a crap result in to a good result. It will just help you squeeze out a bit more performance.

在完成了对不同算法的研究之后，选择一个适用于您的数据集的算法，是时候从中得到最好的结果了。你可以用几种方法来做到这一点，但重要的是要确保你的结果在这一点上是重要的，因为超参数调整不会把废话结果变成一个好结果。它只会帮助你挤出更多的表现。

Here are some standard ways to improve an already working algorithm.

以下是一些常用的标准方法，可以改善已经使用的算法。

**Hyper-parameter Tuning:** All algorithms have hyper-parameters and making sure these are optimal is key to getting the best performance.

超参数调优：所有算法都具有超参数，确保这些参数是最佳的，是获得最佳性能的关键。

**Ensemble Methods:** Where predictions are made by combining multiple models

集合方法：通过组合多个模型进行预测。

**Extreme Feature Engineering:** Attribute decomposition and aggregation seen in data preparation is pushed to the limits

极端特征工程：在数据准备中看到的属性分解和聚合被推到极限。

**Present Results**

**表述结果**

The results of a complex machine learning problem are often meaningless in a vacuum. It’s important to put them in context. This typically means a presentation to stakeholders. This applies to big meetings with CEOs and online competitions. Its good practice and gives everyone involved a good understanding of the problem and how you solved it.

一个复杂的机器学习问题的结果脱离实际，往往毫无意义。把它们放在背景中是很重要的。这通常意味着向利益相关者作表述。比如说，有CEO参加的大型会议和在线竞赛。这是一个非常好的实践，让每一个涉及的人对这个问题有一个好的理解以及你是如何解决这个问题。

Here is a quick template for you to present your results:

以下是表述结果的简洁模板：

**Why:** Define the environment that the problem exists in and set up a motivation for the solution

为什么：定义要解决的麻烦所在的环境，提供解决方案动机。

**Question:** Describe the problem as a question that you went out and answered.

问题：把麻烦描述成你提出的一个问题，并回答之。

**Solution:** Concisely describe the solution as an answer to the question you just posed

解决方案：简要地将解决方案描述为刚刚提出的问题的答案。

**Findings:** List out all of the discoveries you made while solving the problem.

调查结果：列出你在解决问题时所做的所有发现。

**Limitations:** Clearly go over the limitations of the model. What is it not good at and what can be done better.

局限性：清楚地考察模型的局限性。什么不擅长，什么可以做得更好。

**Conclusions:** Go back to the why, question, and solutions and tie it together in a way that makes it easy to remember.

结论：回到原因，问题和解决方案，并以易于记忆的方式将其结合在一起。

**Remember that this is not the end all be all of processes, but it is a good step towards becoming a machine learning engineer.**

请记住，这不是最终的全部过程，但却是成为一名机器学习工程师的一个非常好的一步。