# Entry 1

## The Problem

I am qualified. Quantitively, I know this is true: - I have a Master of Science in Data Science - I have a job on a machine learning team at a large company - I’ve worked my way through multiple courses on Coursera, Udemy, EdX, and Udacity - I’ve started trying my hand at competitions on Kaggle and have worked my way through several of their tutorials - I’ve read blogs and books, listened to podcasts, worked my way through a plethora tutorials, and read more documentation than I’d wish on my worst enemy

Still, I feel like [an impostor](https://towardsdatascience.com/how-to-manage-impostor-syndrome-in-data-science-ad814809f068) who doesn’t know what I’m doing.

So the question is, **how do I improve my skills and increase confidence in my abilities?**

Perhaps it’s time to try a different track.

Perhaps I know too much, am getting bogged down in the edge cases, and need to [refocus on the fundamentals](https://jamesclear.com/fundamentals).

Perhaps I should try walking before I run. Or even crawling before I walk. Or…well…you get the idea.

## The Options

A friend introduced me to Haikyuu, an anime about a bunch of high school boys playing volleyball. It wasn’t until I blew through the anime and was most of the way through the underlying manga that I realized that some, or even most, of the work and strategies athletes use to improve their sports skills can be adapted to improving technical skills.

Having grown up playing sports, I’m no stranger to the hard work that goes into improving physical skills and mental resiliance. I’m also no stranger to the amount of research that goes into making athletes stronger, faster, and more psychologically prepared. So I made a list of strategies that I remembered or could find online that athletes use:

* List from notebook

Sounds like a lot of work. Especially with advice like [Fall in Love with Boredom](https://jamesclear.com/in-love-with-boredom) or the more reasonably titled [Fall in Love with Systems](https://jamesclear.com/goals-systems), both of which offer the same advice: focus on what you’re doing, not where you’re going. And who isn’t familiar with the [10,000 Hour Rule](https://en.wikipedia.org/wiki/Outliers_(book))?

## The Proposed Solution

After making the list, I happened upon the [Machine Learning Mastery website](https://machinelearningmastery.com/youre-wrong-machine-learning-not-hard/) in one of my searches for information. Once I read about what I came for ([L1 and L2 norms](https://machinelearningmastery.com/vector-norms-machine-learning/)), I started exploring the site. He presents a good case for a top-down approach wherein you solve one single small problem at a time and learn as you go instead of a bottom-up approach wherein you learn all the math and all the programming and all the machine learning and all the algorithmic theory before applying it. I decided to try his approach.

One of the recommendations that resonated with me revolved around creating a write up after every project. The goal of this diary is to take the smallest problem I can find and solve just one aspect. Once that aspect has been addressed, I can move on to the next.

The order of the challenges to solve immediately became my first challenge. **How to pick what to work on first**?

## The First Step

#### The Problem

There are many things to address even before looking at algorithms like linear regression, SVM, or k Nearest Neighbors. First the data has to be preprocessed. What to do with missing values? Most algorithms choke if there’s no value for a specific feature of a specific observation. Others, like decision trees, could care less. What about categorical variables. Most algorithms prefer numeric values. The fact that you have seven different kinds of mushroom mean nothing to them. But if you just change them into numbers, then the algorithm thinks button mushrooms come before shiitake.

#### The Proposed Solution

Jason Brownlee makes an eloquent argument for choosing or developing a [systematic process](https://machinelearningmastery.com/process-for-working-through-machine-learning-problems/) to work through a problem. Having spent hours upon hours working through data in a one-off manner, a process sounds like a stellar idea.

Having a clear process gives direction on what to do, when to do it, and what comes next. It also provides a framework for incorporating [deliberate practice](https://jamesclear.com/deliberate-practice-theory) into the pipeline, concentrating on one single aspect to improve. This will allow me to generate the [quantity](https://jamesclear.com/repetitions) of work that should produce better quality as antidoctially suggested in [Art & Fear](https://www.amazon.com/gp/product/0961454733/ref=as_li_qf_sp_asin_il_tl?ie=UTF8&camp=1789&creative=9325&creativeASIN=0961454733&linkCode=as2&tag=jamesclear-20&linkId=CYEZ57AX7IODGHWX), based on the actual results of a photography class at the University of Florida.

## Choosing a process

Jason recommends a [5 step process](https://machinelearningmastery.com/process-for-working-through-machine-learning-problems/) on his Machine Learning Mastery website. This doesn’t seem quite detailed enough for me.

Chapter Two of [Hands on Machine Learning with Scikit-Learn & TensorFlow](https://www.amazon.com/Hands-Machine-Learning-Scikit-Learn-TensorFlow/dp/1491962291) (yes the first edition, because that’s the edition I bought during my graduate studies and I own quite enough data science books already thank you very much) proposes an 8 step process. This process better aligns with my experiences with the requirements for a machine learning solution.

The 8 steps are:

1. Frame the problem and look at the big picture
2. Get the data
3. Explore the data to gain insights
4. Prepare the data to better expose the underlying data patterns to Machine Learning algorithms
5. Explore many different models and short-list the best ones
6. Fine-tune the models and combine them into a great solution
7. Present the solution
8. Launch, monitor, and maintain the system

### Added bonus

The added bonus of using these steps is that each one helps pick as the next challenge to tackle.

Next up, Frame the Problem.