Machine Learning Mastery
Making Developers Awesome at Machine Learning

Search...

The Machine Learning Mastery Method

by **Jason Brownlee** on October 10, 2016 in **Start Machine Learning**

Tweet   Tweet   **Share**   **Share**

5-Steps To Get Started and Get Good at Machine Learning

I teach a 5-step process that you can use to get your start in applied machine learning.

It is unconventional.

The traditional way to teach machine learning is bottom-up.

Start with the theory and math, then algorithm implementations, then send you off to figure out how to start solving real-world problems.
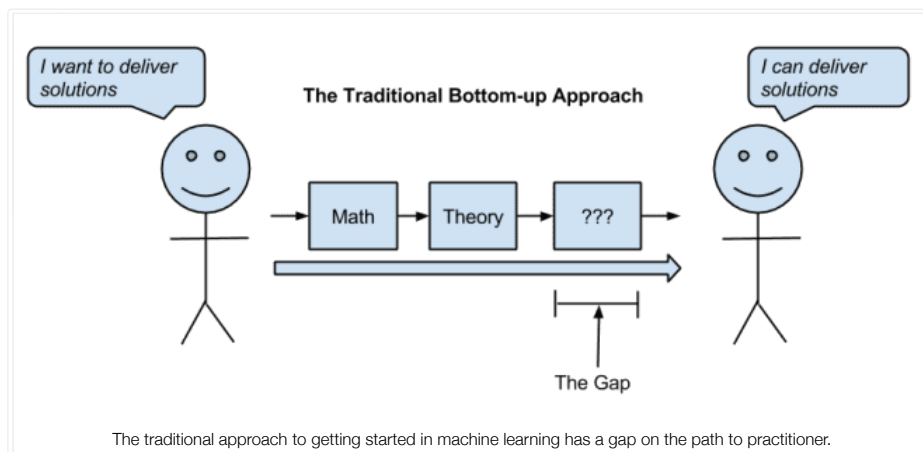
The traditional approach to getting started in machine learning has a gap on the path to practitioner.

The Machine Learning Mastery approach flips this and starts with the outcome that is most valuable.

It targets the outcome that business wants to pay for:
how to deliver a result.

A result in the form of a set of predictions or model that can reliably make predictions.

This is a top-down and results-first approach.

Starting with the goal of achieving the result that is most desirable in the marketplace, what is the shortest path to take you, the practitioner, to that result?

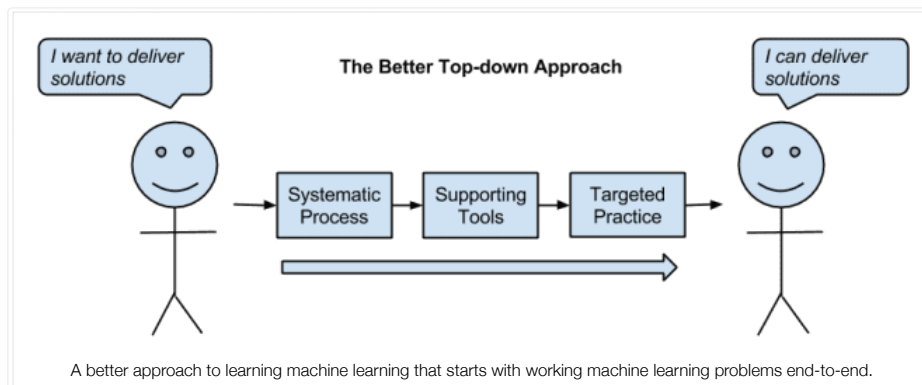We can summarize this path in 5-steps as follows:

- **Step 1: Adjust Mindset** (*believe!*).
- **Step 2: Pick a Process** (how to get results).
- **Step 3: Pick a Tool** (implementation).
- **Step 4: Practice on Datasets** (*put in the work*).
- **Step 5: Build a Portfolio** (*show your skills*).

That's it.

This is the philosophy behind all of my Ebook training.

It's why I created this website. I knew an easier way and just had to share it.

Below is a cartoon to illustrate the process, where step 1 (on mindset) and step 2 (on show your work) are omitted for brevity.

A better approach to learning machine learning that starts with working machine learning problems end-to-end.

Let's take a closer look at each step.

Step 0: Landmarks

Before we begin, you must know the landmarks of machine learning.

I often just assume this, but you cannot proceed unless you know some true basics.

For example:

- You should know what machine learning is and be able to explain it to a colleague.
  - What is Machine Learning?
- You should know some examples of machine learning problems off the top of your head
  - Practical Machine Learning Problems
- You should know that machine learning is the only way to solve some complex problems.
  - Machine Learning Matters
- You should know that predictive modeling is the most useful part of applied machine learning.
  - Gentle Introduction to Predictive Modeling
- You should know where machine learning fits with regard to AI and Data Science
  - Where Does Machine Learning Fit In?
- You should know the types of machine learning algorithms available.
  - A Tour of Machine Learning Algorithms
- You should know some basic machine learning terms
  - How To Talk About Data in Machine Learning

Step 1: Mindset

Machine learning is not just for the professors.

It is not just for the gifted or the academics.

You Must Believe

You can learn the topic and apply it to solve problems.

There's no reason why not.

- You do not need to write code.
- You do not need to know or be good at math.
- You do not need a higher degree.
- You do not need big data.
- You do not need access to a supercomputer.
- You do not need a lot of time.

It is so very easy to come up with excuses to not get started in machine learning.
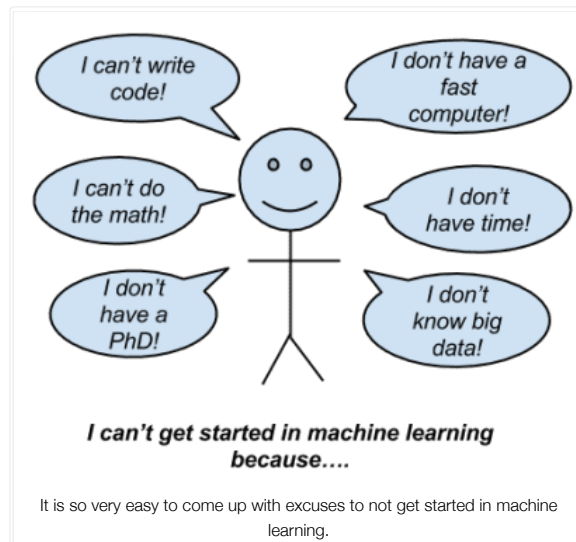
Really, there is only one thing that can stop you from getting started and getting good at machine learning.

It's you.

- Maybe you just can't find the motivation.
- Maybe you think you have to implement everything from scratch.
- Maybe you keep picking advanced problems rather than beginner problems to work on.
- Maybe you don't have a systematic process to follow in order to deliver a result.
- Maybe you're not making use of good tools and libraries.

Clear the limiting beliefs stopping you from getting started.

This post might help:

- What is Holding you Back From Your Machine Learning Goals?

There are a lot of speed bumps you can hit.

Identify them, address them, and keep moving.

Why Machine Learning?

Once you know that you can do machine learning, understand why.

- Maybe you're interested in learning more about machine learning algorithms.
- Maybe you're interested in creating predictions.
- Maybe you're interested in solving complex problems.
- Maybe you're interested in creating smarter software.
- Maybe you're even interested in becoming a data scientist.

Think hard on this topic and try and figure out your "*why*".

This post might help:

- [Why Get Into Machine Learning?](#)

Once you have your "*why*", find your tribe.

Which group of machine learning practitioners do you have the most affinity?

- Maybe you're a business person with a general interest.
- Maybe you're a manager delivering a project.
- Maybe you're a machine learning student.
- Maybe you're a machine learning researcher.
- Maybe you're a researcher with a sticky problem.
- Maybe you want to implement algorithms
- Maybe you need one-off predictions.
- Maybe you need a model you can deploy.
- Maybe you're a data scientist.
- Maybe you're a data analyst.

Each tribe has different interests and will approach the field of machine learning from a different direction.

Not all books and materials are right for you, find your tribe, then find the materials that speak to you.

This post might help:

- [Find Your Machine Learning Tribe](#)

Step 2: Pick a Process

Do you want to reliably get above average results on problem after problem?

You need to follow a systematic process.

- A process allows you to harness and reuse best practices.
- It means you don't have to rely on memory or intuition.
- It guides you through a project end-to-end.
- It means that you always know what to do next.
- It can be tailored to your specific problem types and tools.

A systematic process is the difference between a roller coaster of good and bad results on the one hand and above average and forever improving results on the other.

I would choose above average and forever improving results every time.

A process template that I recommend is as follows:

- **Step 1**: Define your problem.
- **Step 2**: Prepare your data.
- **Step 3**: Spot-check algorithms.

- **Step 4**: Improve results.
- **Step 5**: Present results.

Below is a nice cartoon to summarize this systematic process:

Select a systematic and repeatable process that you can use to deliver results consistently.
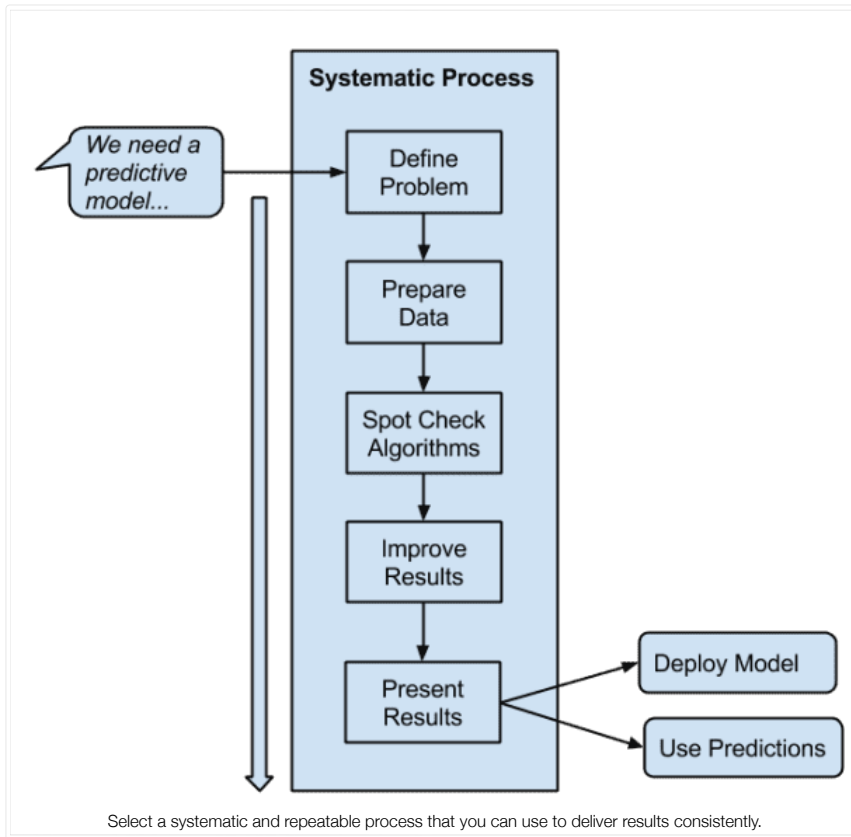
You can learn more about this process in the post:

- Applied Machine Learning Process

You do not have to use this process, but you do need a systematic process for working through predictive modeling problems.

Step 3: Pick a Tool

Pick a best-of-breed tool that you can use to deliver machine learning results.

Map your process onto the tool and learn how to use it most effectively.

There are three tools I recommend the most:

- **Weka Machine Learning Workbench** (*Perfect for beginners*). Weka offers a GUI interface and no code is required. I use it for quick one-off modeling problems.
    - Weka Machine Learning Mini-Course
- **Python Ecosystem** (*Perfect for intermediate*). Specifically pandas and scikit-learn on top of the SciPy platform. You can use the same code and models in development and they are reliable enough to run in operations.
    - Python Machine Learning Mini-Course
- **R Platform** (*Perfect for advanced*). R was designed for statistical computing, and although the language is arcane and some of the packages are poorly documented, it offers the most methods as well as state of the art techniques.
    - R Machine Learning Mini-Course

I also have recommendations for specialty areas:

- **Keras for Deep Learning**. It uses Python meaning you can leverage the whole Python ecosystem which saves a lot of time. The interface is very clean, whilst also supporting the power of the Theano and Keras back-ends.
    - Deep Learning Mini-Course
- **XGBoost for Gradient Boosting**. It is the fastest implementation of the technique around. It also supports both R and Python allowing you to leverage either platform in your project.
    - XGBoost Mini-Course

These are just my personal recommendations and I have lots of posts as well as more detailed training on each.

Learn how to use your chosen tool well. Study it. Become an expert in it.

What Programming Language?

The programming language does not matter.

Even the tool you use does not matter.

The skills you learn working through problems will transfer from platform to platform easily.

Nevertheless, here are some survey results on the most popular languages in machine learning:

- Best Programming Language for Machine Learning

Step 4: Practice on Datasets

Once you have a process and a tool, you need to practice.

You need to practice a lot.

Practice on standard machine learning datasets.

- Use real-world datasets, collected from an actual problem domain (rather than contrived).
- Use small datasets that fit into memory or an excel spreadsheet.
- Use well-understood datasets so you know what kind of results to expect.

Practice on different types of datasets. Practice on problems that make you uncomfortable as you will have to push your skills to get a solution. Seek out different traits in data problems, such as:

- Different types of supervised learning such as classification and regression.
- Different sized datasets from tens, hundreds, thousands and millions of instances.
- Different numbers of attributes from less than ten, tens, hundreds and thousands of attributes.
- Different attribute types from real, integer, categorical, ordinal and mixtures.
- Different domains that force you to quickly understand and characterize a new problem in which you have no previous experience.

Use the UCI Machine Learning Repository

These are the most used and best-understood datasets and the best place to start.

Learn more in the post:

- Practice Machine Learning with Small In-Memory Datasets from the UCI Machine Learning Repository

Use machine learning competitions, such as Kaggle

These datasets are often larger and require more preparation to model well.

For a list of the most popular datasets that you could practice on, see the post:

- Tour of Real-World Machine Learning Problems

Practice on problems of your own devising

Collect data on machine learning problems that matter to you.

You will find the problems and the solutions you devise so much more rewarding.

For more information, see the post:

- Work on Machine Learning Problems That Matter To You

Step 5: Build a Portfolio

You will build up a collection of completed projects.

Put them to good use.

As you work through datasets and get better, create semi-formal outputs that summarize your findings.

- Maybe upload your code and summarize it in a readme.
- Maybe you write up your results in a blog post.
- Maybe you make a slide deck.
- Maybe you create a little video on youtube.

Each one of these completed projects represents one piece of your growing portfolio.

Just like a painter, you can build a portfolio of completed work to demonstrate your growing skills in delivering results with machine learning.

You can learn more about this approach in the post:

- Build a Machine Learning Portfolio

You can use this portfolio yourself, leveraging code and knowledge in your prior results in larger and more ambitious projects.

Once your portfolio is mature, you may even choose to leverage it into more responsibility at work or into a new machine learning focused role.

For more on this see the post:

- Get Paid To Apply Machine Learning

Tips And Tricks

Below are some practical tips and tricks you may consider when using this process.

- Start with a simple process (*like above*) and a simple tool (*like Weka*), then advance once you have confidence.
- Begin with the simplest and most used datasets (*iris flowers* and *Pima diabetes*).
- Each time you apply the process, look for ways to improve it and your usage of it.
- If you discover new methods, figure out the best way to integrate them into your process.
- Study algorithms, but only as much and in ways that help you achieve better results with your process.
- Study and learn from experts and see what methods you can steal and add to your process.
- Study your tool like you do predictive modeling problems and get the most out of it.
- Tackle harder and harder problems, leave the easy ones as you won't learn much from them.
- Focus on clearly presenting results, the better you do this, the greater the impact of your portfolio.
- Engage in the community on forums and Q&A sites, both ask and answer questions.

Summary

In this post, you discovered a simple 5-step process that you can use to get started and make progress in applied machine learning.

Although simple to layout, the approach does take hard work, but it does payoff.

Many of my students worked through this process and got work as machine learning engineers and data scientists.

If you are in a deeper treatment of this process and related ideas, see the post:

- Machine Learning for Programmers

Do you have any questions?
Ask in the comments below and I will do my best to answer.
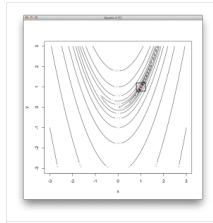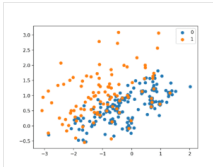
Tweet   Tweet   Share   Share

More On This Topic

Convex Optimization in R

14 Different Types of Learning in Machine Learning

11 Classical Time Series Forecasting Methods in…

Undersampling Algorithms for Imbalanced Classification

Machine Learning for Developers

Essence of Bootstrap Aggregation Ensembles

About Jason Brownlee

Jason Brownlee, PhD is a machine learning specialist who teaches developers how to get results with modern machine learning methods via hands-on tutorials.

View all posts by Jason Brownlee →

42 Responses to *The Machine Learning Mastery Method*

Mrutyunjaya October 10, 2016 at 3:20 pm #

REPLY ↩

Hi Jason, Thanks for sharing. A great content for beginners.

REPLY ↩