

Get started

Open in app



501K Followers · About Follow

You have **2** free member-only stories left this month. [Sign up for Medium and get an extra one](#)

OPINION

Want to Be a Data Scientist? Don't Start With Machine Learning.

The biggest misconception aspiring data scientists have



Terence S · 1 day ago · 5 min read ★



Photo by [Will Porada](#) on [Unsplash](#)

The first thing most people think about when they hear the term “data science” is usually “machine learning”.



looking for a place to start learning about data science, you can guess where I started (hint: it rhymes with *bean churning*).

This was my biggest mistake and this leads me to my main point:

If you want to be a data scientist, don't start with machine learning.

Bear with me here. Obviously, to be a “complete” data scientist, you'll have to eventually learn about machine learning concepts. But you'd be surprised at how far you can get without it.

So why shouldn't you start with machine learning?

1. Machine learning is only one part of a data scientist (and a very small part too).

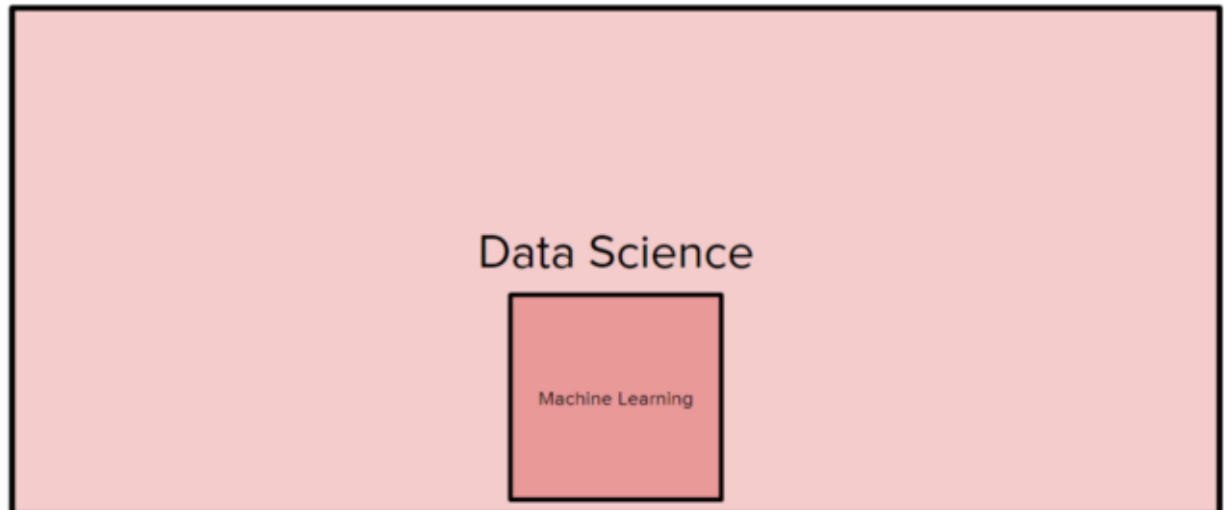


Image created by Author

Data science and machine learning are like a square and a rectangle. Machine learning is (a part of) data science but data science isn't necessarily machine learning, similar to how a square is a rectangle but a rectangle isn't necessarily a square.

In reality, I'd say that machine learning modeling only makes up around 5–10% of a data scientist's job, where most of one's time is spent elsewhere, which I'll elaborate on later.



2. Fully understanding machine learning requires preliminary knowledge in several other subjects first.

At its core, machine learning is built on statistics, mathematics, and probability. The same way that you first learn about English grammar, figurative language, and so forth to write a good essay, you have to have these building blocks set in stone before you can learn machine learning.

To give some examples:

- Linear regression, the first “machine learning algorithm” that most bootcamps teach first is really a **statistical** method.
- Principal Component Analysis is only possible with the ideas of matrices and eigenvectors (**linear algebra**)
- Naive Bayes is a machine learning model that is completely based on Bayes Theorem (**probability**).

And so, I’ll conclude with two points. One, learning the fundamentals will make learning more advanced topics easier. Two, by learning the fundamentals, you will already have learned several machine learning concepts.

3. Machine learning is not the answer to every data scientist’s problem.

Many data scientists struggle with this, even myself. Similar to my initial point, most data scientists think that “data science” and “machine learning” go hand in hand. And so, when faced with a problem, the very first solution that they consider is a machine learning model.

But not every “data science” problem requires a machine learning model.

In some cases, a simple analysis with Excel or Pandas is more than enough to solve the problem at hand.

In other cases, the problem will be completely unrelated to machine learning. You may be required to clean and manipulate data using scripts, build data pipelines, or create interactive dashboards, all of which do not require machine learning.

What should you do instead?



fundamentals. And I still stand by this.

Like I said before, learning the fundamentals will make learning more advanced topics easier, and by learning the fundamentals, you will already have learned several machine learning concepts.

I know it may feel like you're not progressing to be a "data scientist" if you're learning statistics, math, or programming fundamentals, but learning these fundamentals will only accelerate your learnings in the future.

You have to learn to walk before you can run.

If you would like some tangible next steps to start with instead, here are a couple:

1. **Start with statistics.** Of the three building blocks, I think statistics is the most important. And if you dread statistics, data science probably isn't for you. I'd check out Georgia Tech's course called [Statistical Methods](#), or [Khan Academy's video series](#).
2. **Learn Python and SQL.** If you're more of an R kind of guy, go for it. I've personally never worked with R so I have no opinion on it. The better you are at Python and SQL, the easier your life will be when it comes to data collection, manipulation, and implementation. I would also be familiar with Python libraries like Pandas, NumPy, and Scikit-learn. I also recommend that you learn about binary trees, as it serves as the basis for many advanced machine learning algorithms like XGBoost.
3. **Learn linear algebra fundamentals.** Linear algebra becomes extremely important when you work with anything related to matrices. This is common in recommendation systems and deep learning applications. If these sound like things that you'll want to learn about in the future, don't skip this step.
4. **Learn data manipulation.** This makes up at least 50% of a data scientist's job. More specifically, learn more about feature engineering, exploratory data analysis, and data preparation.

Get started

Open in app



machine learning should not be the focus of your studies because it's not a great use of time and will do little in helping you be a successful data scientist in the working world.

With that said, I wish you the best of luck in your endeavors!

Not sure what to read next? I've picked another article for you:

How I'd Learn Data Science if I Could Start Over (2 years in)

My guide to learning data science as efficiently as possible

towardsdatascience.com

Terence Shin

- *If you enjoyed this, [follow me on Medium](#) for more*
- *Sign up for my email list [here](#)!*
- *Let's connect on [LinkedIn](#)*
- *Interested in collaborating? Check out my [website](#).*

Sign up for The Daily Pick

By Towards Data Science

Hands-on real-world examples, research, tutorials, and cutting-edge techniques delivered Monday to Thursday. Make learning your daily ritual. [Take a look](#)

Your email

Get this newsletter

By signing up, you will create a Medium account if you don't already have one. Review our [Privacy Policy](#) for more information about our privacy practices.

Get started

Open in app



[About](#) [Help](#) [Legal](#)

Get the Medium app

