**The Mindset of a Successful Data Scientist**

Four Attitudes for Success in Data Science

There are so many articles out there about how to become a great data scientist and they are almost entirely focused on technical skillsets or processes. But equally, if not more, important is the data scientist’s mindset. I see it time and time again — a smart person with all the right technical skills fails to succeed in a data science role because their analytical mindset is all wrong.

They know their models, they know their stats, they’re strong coders but they struggle to take a problem and find a solution. Their mind is overly focused on the technical. They miss the question for the solution.

So much data science culture is built around technical skills. “What are the techniques and libraries you know?” But Data Science teams are tasked with solving business problems and business problems require creative problem solving. Creative problem solving that can only be achieved with the right mindset.

What is the mindset of a successful data scientist?

***Seek out the right tool***

In a data science role you have to understand that your team’s end goal is **always** to deliver answers to business problems. It is **never** to implement a specific type of model or overcome a specific technical hurdle.

Goals of implementing specific model types or overcoming specific technical hurdles are secondary to answering the client’s biggest questions. They are the steps along the way to get there, but they are not the end goal.

Successful data scientists seek out the big question. **Then** they find the right technique to address the question. As a data scientist you should never look at your project as “I need to implement a neural network”, but as “I need to solve this classification problem and a neural network seems like the most suitable technique right now.” There’s always a possibility you might find a more suitable technique along the way.  
  
To put it another way, it is always about the solution and it is never about the tool. No matter what anyone tells you, your end client cares about solutions, not tools. Your ability to force a tool into a solution will not be appreciated if that tool wasn’t the best one for the job to begin with.

So, what is a big question? A big question is about what the end client is going to do with the model, solution, or insights you provide them. What are they trying to learn and how are they going to use those insights? For example, “I need to understand which of our subscribers is an active runner because I intend to send those subscribers a targeted e-mail.” Which would lead to follow-up questions such as “how bad would it be to send the offer to a non-active runner?” In data science speak: “are false positives ok?” The answers to those follow-ups would help us to fine-tune the model and might even cause us to change course on the type of model implemented.

You might find that your leaders aren’t giving you this information. That the question has been distilled to a specific technical task when it lands on your desk. You must seek it out! Your leader has likely used their experience to simplify the task, and that is valuable. But if they haven’t informed you, then you must ask “what is the client going to do with this? In what situation should I begin to consider alternatives?” Not only will this enable your success on the specific project, but it will accelerate your learning as you absorb your leader’s experience.

Always, in all situations, ask yourself “what am I really trying to solve?” Only then can you ask yourself “what is the right tool for this job?”

***Respect stakeholders***

So now you know what the big question is and you’ve spent time reflecting on the right tool for the job. It’s time to go implement it and get some answers! Right? But what about all the things your leader, your client, and your colleagues already know?

How often have you heard the tale of artificial intelligence models producing outlandish predictions that were just *obviously* wrong? IBM’s Watson competing on Jeopardy is a great example. It was a huge success; it outperformed the best human competitors. At the same time, it produced some wildly wrong answers. Any human, with any context at all, knew those answers were wrong. It was laughable from our perspective. But from the machine’s perspective, it was a perfectly valid response. The machine was just missing a piece of context that humans are good at adding in.  
  
As a data scientist you are trying to solve problems by digging into the data, the math, and the logic. All of that exists in a wider context. If you don’t understand the wider context you will produce laughable results. Your clients, your leader, and your colleagues will notice something is wildly wrong and if you can’t explain it your solution will be doomed. It’s unfortunate because your model might be the best one that exists, it might outperform all the best competing models. But those laughably wrong results, however small, will make it look like a fool.

When someone asks a question, there is often a huge wealth of auxiliary contextual knowledge to draw on. If they’re asking about active runners because they want to introduce a targeted e-mail for them, they probably know something about active runners. At best their knowledge might be useful in training the model. At worst, they’ll be irrelevant for our purposes. Most likely, they will help prepare us for questions when the client sees the end results. “Hey, your classification shows 25% of our customers are active runners, but we thought it was 35%? What happened?”

So often Data Scientists will neglect that knowledge — “that’s not what my data set says” or worse “who cares what they *think*, I have data and it *shows* ….” That’s akin to IBM saying “it doesn’t matter that all you humans could see our answer was wrong, our answer was just better than the truth.”

You must do your best to learn the wider context at the beginning of the project. Some of it will come from your clients, some of it will come from your leader, and some of it will come from your colleagues. Consider the active runner example above. Your first question might be: “Do we have any existing estimates of what portion of our customers are active runners? How confident are we in those estimates? Where did they come from?” This will give you a strong base to judge whether your model is systematically under or overestimating the classification. You might have to talk to all three groups of stakeholders before you get answers, but you will save yourself a lot of re-work and embarrassment.

Context matters! Successful data scientists find it and use it. Those that don’t wind up looking foolish — no matter how great their actual work.

***Know how to be wrong***

You’ve fully prepared, you know the question, you know the context, you’ve found the best technique and you’ve fine tuned your model. You’ve passed your model to the next team. But something, somewhere, hasn’t worked and your results, your theory, and your model are wrong.  
  
However rare, it happens to all of us. Sometimes it’s a simple error and it’s easily fixed, sometimes it’s a deep systematic error that takes weeks to correct. Finding success means building on these errors in a positive way. You must be wrong in the right way.  
  
When you realize you’re wrong it’s a huge shock to the system. A lot of emotion gets involved very quickly. You feel embarrassment, discouragement, and anger. For many people that means they ignore the fact that they, or their model, could be wrong. It’s easier to neglect the situation altogether than it is to deal with all that emotion. The first rule to being wrong is that you must be open to the possibility that something is wrong. No matter how bad the initial error, doubling down on something that is false always increases the problems.  
  
So what is the successful approach? Listen to others when they say you’re wrong. Be open to their input. If it is too late in the process to be wrong and you can feel the emotions coming on, then stop — immediately. You first need to know if you’re actually wrong. The only way to do so is to step back from the situation and understand the feedback you’re being given. If it means you must move on to a new topic, or cancel a meeting altogether, then do it.

As a data scientist the truth is the most important goal — not supporting your ego. So, ask for time to evaluate the situation, and if you have to eat humble pie later, then eat it. It might hurt in the short-term but in the long run you will build relationships, trust and collaboration by being open to opposing points of view. In the short-term you’ll also protect an error from cascading into larger and larger problems as they so often do when hidden.

Always be open to the possibility of errors and always choose the truth over ego.

***Love solving problems***

The most important thought for a Data Scientist is “I can’t wait to get to the bottom of this problem.” As a Data Scientist the variety and complexity of questions can be overwhelming. If you’re not thinking “there must be a way to get to an answer, I just need to find it” then things will quickly get overwhelming. The key here is that you have to be thinking “there’s a problem and I *will*find a solution.”  
  
On top of that, your clients and your leader are looking for answers and, unfortunately, the road to those answers is rarely the smooth straight line that so many new data scientists expect. In truth, it is often a long and winding road. There will be many missteps and dead ends. Face them with an attitude of opportunity — they are learning experiences, however frustrating they may be at the time, and they will serve you well in the future.

You must love the process of turning a question into an answer and along the way you must love the process of overcoming each individual technical hurdle. Your client doesn’t care about those technical hurdles — but you need to love the process of beating them. Those technical hurdles will define your day-to-day experience; you need to feel the joy in overcoming each of them.

Let yourself feel the big satisfaction when you have an answer, celebrate it and share it. But don’t lose sight of the joy in the journey.

While technical skills might get you in the door, those skills won’t lead to success unless combined with the right mindset. To build a career in data science you must constantly ensure you maintain the right mindset — seek out the right tool, listen to your stakeholders, know how to be wrong and, most importantly, be passionate not only about the destination but the journey along the way.