Forecasting

# Avoid Wasting Time

### Always spend time sizing the problem and getting an early read for feasibility if you can

That is, build the simplest model you can with the best data you have and see how things look. My litmus test historically has been that if I build a relatively simple/naive model and my quality of fit is <0.1 then odds are this is never going to become a 0.8+ model. I may get it up to the .2s, maybe 0.3-0.4, but if a basic model gets you nowhere near the reservation, it's highly unlikely that through sheer modeling you will close that big of a gap.

### Always think through the whole process - and I do mean the whole process

Where is the data going to come from, when is the data going to be available in reality/database/your environment, how quickly can you get the data, how much data will there be, how clean will future data be, how many people need to be involved in getting you the data.

You run your model - can you scale it? How fast does it need to run? How many times does it need to run? How many data does it need to write? How big of an environment will you need?

Downstream: your model produces an output. How does that propagate through systems? processes? Who needs to use the output? What are they going to do? Will they get the output so that the timing allows them to make decisions? How good does your output need to be for the decisions to be improved?

### Don't be afraid to push people who are a) bad at giving clear problem statements, or b) bad at thinking through how complex what they are asking actually is

This is something that took me a while to learn. If someone tells me "yeah, we need a model that predicts the best price", in the past I would have gone off on my own and figured out what that meant. I've learned that you need to sit there and ask questions even though the other person may get annoyed by it. This is how these conversations normally go:

Me: "What do you mean by best price?"

Them: "You know, the best one. The one that makes us the most money".

Me: "More money meaning revenue? Or profit?"

Them: "Both".

Me: "We can't do that - we have high profit items that are low revenue and vice versa. We need to pick one or the other".

Them: "But I want both, I want us to make the most money! It's not that complicated!"

Me: "It's not complicated, but you can either maximize revenue or profit. They are not the same. My suggestion would be to maximize profit - that is the amount of money that we actually make as a company"

Them: "Ok, fine, just as long as it makes us the most money"

Me: "Ok, sounds good, so we agree that the goal is maximize profit?"

Them, reluctantly: "Yeah, i guess".

Follow up with an email saying "as per our meeting, we will be building this model with the objective to maximize profit".

Don't get me wrong, these conversations suck, because the person on the other end can act like a child, But you're better off having it early and having one uncomfortable conversation than having to have a bunch of them after things don't go well.

### Source

https://www.reddit.com/r/datascience/comments/c2xxt4/has\_anyone\_ever\_wasted\_months\_on\_a\_project/ernsat9