So You Want a Job: Interviewing Data Miners and Modelers

By Tim Graettinger

Judging by the recent posts on numerous web sites, many companies are looking for people with data mining and modeling (DM) skills. I recently interviewed some candidates for an open position. Since people in this field are a unique group, and interviewing them can be challenging, I thought you might like to hear about my approach.

Philosophy for Hiring and Interviewing

When I interview a DM candidate, let's call her Janet, I want to find out:

- Is she smart?
- Is she focused/does she get things done?
- Is she compatible/can I work with her?

Rolled into these three general criteria are several, integrated elements that deserve more elaboration. To me, in a DM context, smart means problem-solving smart. I want to know if Janet is a problem solver. Can she frame a problem? How does she think? Does she believe she can and will get smarter and improve her thinking and skills over time? I believe that intelligence and creativity can be developed and enhanced over time. I want to work with people who feel the same way.

Next, focus and a getting-things-done mentality usually manifest themselves in project work. In the interview with Janet, I want to learn if she can work on projects independently and also with a team. Does she have a process and the focus to stay on track and see a project through to the finish line? How does she work through the inevitable problems that crop up?

Last, I want to know what a working relationship will feel like. I'm particularly interested in her ability to communicate difficult concepts, issues, data, models, and results. And I think those communication skills (or lack thereof) come out clearly in discussing recent projects, as we will see below.

That's the philosophy: Find someone, maybe Janet, who is smart, focused, and compatible. Now, let's get specific. What does the interview itself look and feel like? I structure a 1-hour interview into five segments:

- Introductions
- Recent Project
- Modeling Process
- Troubleshooting
- Wrap-Up

Let's discuss each of the five segments in more detail below.

Part 1 – Introductions

The first purpose of my introduction phase is to put Janet at ease. Simple small talk will suffice. I think it's important to create a relaxed atmosphere where Janet and I can speak comfortably.

Next, I outline the structure of the interview to her, so that she understands where we're going and why. I also ask if she has any questions before we get started. From the candidate's perspective, I think this is a very important component. Janet may have a couple burning questions to ask me, and she may feel some anxiety about when, or even whether, she'll get to ask them. I listen to her questions and write them down, <u>on</u> <u>paper</u>. If any of them can be answered quickly, I do so immediately. For the others, I indicate that we'll return to them during the last 10 minutes of the interview.

To wrap up the introduction phase, I tell Janet what my role is, what my responsibilities are, and how I anticipate we might work together. I ask Janet to tell me a little bit about herself. As she responds, I'm primarily listening to her communication skills, more than to the content. Does she express herself well, thoughtfully, and in complete sentences on a topic she knows intimately – herself. If that's a struggle, I make a mental note of it. If there's not much enthusiasm there, I note that, too. Why should I be enthusiastic about her as a candidate if she's not enthusiastic herself?

Part 2 – Tell Me About a Recent Data Mining/Modeling Project

Continuing into the next phase of the interview, I ask Janet to tell me about a recent project she worked on. To me, this is the most important part of the interview, and it's the longest portion, time-wise. After many years working in this field, I have come to appreciate how important it is to be able to <u>tell</u> <u>the</u> <u>story</u> of a project – whether the project is complete or still on-going. As Janet begins to tell the story, I make notes:

- Does she capture my interest at the start?
- Where does she begin? With the results? With the business problem to be solved? Somewhere else?
- Who are the characters and how does she describe them? Is the client a villain or the village idiot? Does she have positive things to say about her team? Is she the hero?
- Is the story disjointed and confusing, or is it presented clearly, in a compelling way?

All along, I ask Janet questions to help me understand the story and the thought process behind it. Typical questions I like to ask are:

- How did you frame the problem? Why?
- What was the unit of analysis? Why?
- What was your favorite part of the project? Why?
- What was your least favorite part of the project? Why?
- What was the most challenging part of the project? Why?
- How long did it take?
- How did you communicate your findings? Who was the toughest person to sell? How did you convince her?
- What did you learn?

Please notice a few things about the queries above. First, I ask, "Why?" a lot. Why, you ask? I find it gives me the most insight into my criteria (smart, focused, and compatible). Further, the "why" questions are the ones clients seem to ask most. A DM has to be able to support the decisions she makes, and there are a multitude of decisions to make in the course of a project.

Second, I ask Janet the trio of questions (favorite part, least favorite part, most challenging part) to assess her "fit" with the open position that I'm trying to fill. If the least favorite part of the recent project is the <u>primary</u> part of the open position, then the fit is not good. On the other hand, if

Janet really liked presenting results to the client or communicating issues across the team, then I start to get excited.

Finally, the last question above, "What did you learn?" is crucial to me, personally, and it often tells me volumes about the candidate. It's especially important in DM since the field is rapidly changing, and there are few "cookie cutter" projects that can be done by rote. A slow response from Janet of "what … did … I … learn … ?" indicates that she hasn't really thought about it before now. I'm looking for people who are continually asking themselves this question. Some people look at every interaction or project as a performance, a test of their intelligence. Others look at those same interactions as opportunities to learn, to engage, to gain more knowledge. I want to hire the latter.

Part 3: Tell Me About Your Modeling Process – In 5 Minutes or Less

The discussion of a recent project naturally flows into my next topic for Janet, "Tell me about your mining/modeling process". This can certainly be a wide-ranging subject, but it's also one that prospective clients routinely ask about – and one that needs a succinct, satisfying response. If Janet references one of the industry-standard processes, like CRISP-DM or SEMMA[1], that's a good sign for me. It indicates that she has thought about the process and sought out some guidance, or that she actually used the process at her previous workplace. Too often, I meet DM's with a seat-of-the-pants attitude towards modeling and the modeling process. It's great to be creative and to look at problems with a fresh perspective. But, that creativity can be layered on top of a solid, baseline process – a process that insures that all the mundane elements are covered well.
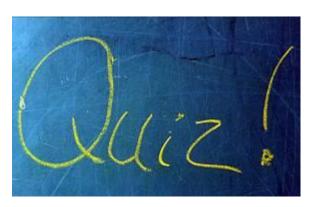

I also ask Janet to draw me a picture (on a whiteboard or on paper) of her process[2]. This is another good opportunity to assess communication skills. As we discuss the picture, I drill in to ask, "Where are your key 'pauses' in the process?" Because good modeling requires reflection, pauses for review – by yourself and with your team - are critical. I want Janet to tell me where she comes up for air, and why.

To round out this portion of the interview, I like to drill down and ask a probing question or two about specific portions of the process. For instance, I ask Janet, "How do you winnow a large number of data elements down to a 'short list' of candidates for modeling?" I prefer questions like this, where there is not a single, universally-accepted method or technique. It gives Janet the opportunity to take a stance, express an opinion, and defend it based on her experience. I also want to hear if she has experimented with different approaches and what she has learned[3] from those experiments.

Part 4: Troubleshooting

In the past, when I worked at software product companies, we often gave candidates a software "coding test". We asked the candidate to write a small function, say, to reverse a linked list. It is hard to design an equivalent, self-contained test for a DM candidate. My approach is to present Janet with a couple troubleshooting problems. I use printed excerpts of report output from a modeling tool. I hand Janet the report and give her some background on the project/application, the data, and the routine that generated the report. I ask her to interpret the report for me, diagnose the problem (if she thinks one exists), and prescribe a plan of action to resolve it.

Examples of troubleshooting exercises I use include:
- A report indicating a mismatch in model performance between training data and testing data
- Regression output containing some elements that "go the wrong way", that is, their impact on the predicted outcome does not match the conventional wisdom and/or other analysis.

From this exercise, I get to see if Janet is smart (does she understand the issue), focused (does she explore various avenues and then follow the most promising one), and compatible (does she engage me and ask me questions, or does she try to resolve the problem without any help). DM's often have to troubleshoot their work or to review the work of colleagues. The interview is the place to uncover Janet's skills for critically examining her own work and the work of others.

Part 5: Wrap-Up

The conclusion of the interview is the time to return to any unanswered questions that were deferred when we started. I also ask Janet if she has any other questions or thoughts that were spurred by our discussion.

At this time, I also give Janet a pitch on why she should want to join the company. If she is a good candidate, she probably has multiple opportunities from which to choose. I need to convince her that ours is the best one to accept. Finally, I thank Janet for coming, and I walk her to her next meeting.

It's my normal practice to immediately return to my desk and write down my notes about Janet. My memories and impressions from the interview fade really quickly, so the sooner I commit them to paper, the better. I finish with a simple "thumbs-up" or "thumbs-down" rating, along with my main reasons why. If I'm unsure, I have to give her my default thumbs-down rating. Spending more time right now to hire the right person more than compensates for the time and aggravation of hiring and working with the wrong person and then letting them go six months or a year later.

Thanks for Coming In

Interviewing and hiring people can be a chore. Or, it can be an opportunity to meet interesting people and even learn something about what they do, how they work and solve problems, and how they communicate. In this article, I described my goals for interviewing (to find out if a candidate is smart, focused, and compatible) and my process (introduction, recent project, DM process, troubleshooting, and wrap-up). I hope this helps you the next time you're face-to-face with a DM candidate.

---

[1] CRISP-DM and SEMMA are two of many process models for data mining that have the industry stamp of approval.

[2] Janet can really impress me if she asks to draw a picture on the whiteboard before I ask her.

[3] There's that theme of learning again.