How I Get My First Job as A Data Scientist

Yufeng (Louis) Wang

Data Scientist at Capital Group | Machine Learning, Relevance, NLP

Hello guys, I am Louis and just graduated from M.S in Business Analytics program of Emory University, Goizueta Business School last month. Luckily, I become a data scientist after graduation. For my undergrad, I studied E-Commerce with a minor in GIS, and had several internships in the area of data science. The job-hunting process is painful, time-consuming, but meaningful. In the last four months, I submitted nearly 500 job applications, had about 90 interviews, and received three offers of data scientist. The purpose of this article is to share some lessons that I learned in this time: How to earn an interview by networking, how to prepare data scientist tech interview, and how to adjust yourself under pressure. I hope this can help people who want to be a data scientist and are still looking for the analytics jobs.

Networking

First of all, from my experience, networking is the most effective way to get an interview. Whether it is a career fair, tech talk, data science conference, info session, or it is a guest speak, be positive to reach out to others. **The key of networking is to find common interests, build connection, and prepare questions.** Do remember to make research about the people you are going to talk with. Apart from the face-to-face communication, LinkedIn is a more powerful tool if you use it correctly. My method of using LinkedIn is (1) to find the target industries and companies, (2) to reach out to the right people (Alumni, friends, ex-colleague... yes, this is the connection!), (3) to send request to connect in LinkedIn (do remember to add a note), (4) to express your interest in the company and position, and (5) to ask for a phone call or referral. By doing this, you can not only know more about the data scientist role, the career suggestions, the detailed background about the industries, but also how to prepare for the interviews. These connections that you accumulate at this time may also be helpful in the future.

Interviews

The title, data scientist, in different companies have different job responsibilities, and the requirements of the skill sets will also be different. To cover most of the common tech questions, I recommend the coming data scientists to prepare four parts: (1) data manipulation using SQL, Python and R, (2) AB testing/Hypothesis testing, (3) machine learning, and (4) case questions. Here I list some useful materials that I used in my job-hunting process.

1.Data manipulation

As a data scientist, you will use nearly 70% of your time on data cleaning. So, the ability of getting the right data to fit your models is super important. You should feel comfortable when

using SQL, Python or R to manipulate the datasets. One of the most common questions for the tech interviews is SQL questions. You should be familiar with windows functions, different between having and where, difference among four types of 'joins', etc. Sometimes the interviewers may first ask you what kind of data you want to use for the specific situations, and why.

Leetcode: https://leetcode.com/problemset/database/

Stanford: https://lagunita.stanford.edu/courses/Home/Databases/Engineering/about

Mode Analytics: https://community.modeanalytics.com/sql/

W3Schools: https://www.w3schools.com/sql/

Write better queries: https://www.datacamp.com/community/tutorials/sql-tutorial-query

Windows functions: http://www.windowfunctions.com/

2. AB testing/Hypothesis testing

One main daily work for data scientists is to find the reason of some phenomenon, and then provide solutions to drive the business issues. So, the hypothesis-driven idea is an important mindset to guide you to the correct direction. You can use the correlation analysis to find the correlation between different variables. You can A/B testing to find the causality of two things. Hypothesis testing is also a must-have skill for data scientists, and it is usually tested in tech companies such as LinkedIn, Facebook, Google, etc. The interviewers may ask you whether to launch a new feature or not, and they will ask you some follow up questions such as how to calculate sample size, what is p-value, how to define the metrics, what is power analysis, what is the expected business impact.

Experiments at Airbnb: https://medium.com/airbnb-engineering/experiments-at-airbnb-e2db3abf39e7

A/B testing fundamentals: https://classroom.udacity.com/courses/ud979

A clear picture of power and significance in A/B tests: http://www.win-vector.com/blog/2014/05/a-clear-picture-of-power-and-significance-in-ab-tests/

AB testing: http://rpubs.com/superseer/ab_testing

PSU statistics: https://onlinecourses.science.psu.edu/statprogram/programs/

3. Machine Learning

Data is the new oil, and machine learning enhance our ability to utilize this new oil. Better understanding concepts and assumptions of different ML models is so important. In most

interviews, I was asked this kind of questions, some from direction questions like what's the tradeoff between bias and variance, what's the assumptions for linear regression, decision tree and logistic regression which one is more robust to data and why, some from a case/product questions like how to improve engagement for facebook, with followed questions like how do you define engagement, what data you want to use, which model you choose and why, how to measure model's performance, etc.

Machine learning Yearning: http://www.mlyearning.org/

Stanford Statistical Learning:
https://lagunita.stanford.edu/courses/HumanitiesSciences/StatLearning/Winter2016/about

Machine learning techniques:
https://www.youtube.com/playlist?list=PLXVfgk9fNX2IQOYPmqjqWsNUFl2kpk1U2&disabl
e_polymer=true

CMU machine learning: http://www.cs.cmu.edu/~tom/10701_sp11/lectures.shtml

BOOKS:

- Introduction to Statistical Learning: http://www-bcf.usc.edu/~gareth/ISL/ISLR%20First%20Printing.pdf
- The Elements of Statistical Learning: https://web.stanford.edu/~hastie/Papers/ESLII.pdf

4. Case studies

Analytics is not only a process but also a mindset. It is a set of process of collecting data, cleaning data, building models and using that to improve something. It is also a mindset of how you solve the problem. Analytical thinking is about asking right questions, find right metrics that can drive the business. This part is a little tricky, because it is used to test candidate's mindset, thinking process, business sense, problem-solving skills, and the understanding about the technology applications in business. For people who don't have enough hands-on industrial experience, they may not have a clear idea of what to do when facing the problems. However, it doesn't mean that fresh grad cannot answer these kinds of questions well and become a data scientist. For me, I often refer to my Alumni in tech companies asking what they are doing every day. Reading blogs is also another way to have industrial knowledge. In my free time, I usually look through some interesting blogs in Medium.com and Reddit.com. If you are living in some big cities, do remember to join the data science Meetup group. Apart from that, I also suggest you to read three books to prepare for the case interviews: Case in Points, A collection of data science take-home challenges, and Cracking the PM Interview.

I truly understand that the road to success is paved of failures, because I know it is so hard for a non-experienced international student to find the data scientist job in USA. But thanks to the rejections from those companies, it helps my better understand myself. Reach further, try harder, fail harder and behave better next time. My boss once told me, if you don't fail you are not trying. Failure in life is not only inevitable but also necessary. So, don't feel disappointment and depressed when you get rejection letters. It doesn't mean you are not good enough; it only means you are not a good fit at this time but it can affect nothing about you in the future.

Here is all I want to share with you today. I appreciate everyone who helps me in my job searching process. It is my great luck to start off my career at Best Buy, and I know when I start to work in the real business world, the real learning just begins...