B Q

Tuesday, August 07, 2018

12:44 PM

* + Interest of the company/team :
  + Interest + ability

Learn a lot; interest in DS; not want to focus; explore other;  risk help save money, marketing make more money

Hear from your team members; advanced not conventional method; excited

Know more about the customer; Now more about macroeconomic and default

A lot of similarities;

**Marketing/DS:**

I’ve learned a great deal during my current job in both technical skills like programming, building the models, and also communication skills. It has been a great experience.

My interest has always been the application of advanced data science techniques in finance/banking fields. But I don't want to just focus on one limited fields like the credit risk modeling I am currently doing. *(Especially the current situation of****de-regulation****, there is not too much new thing to learn in the near future.)* So I would like to **try some other fields** in banking.

And I know many of your team members, and constantly hear them discussing some of those fancy machine learning techniques that you guys are using to target customers. So the opportunities to work on different projects using most state-of-the-art data science techniques to get insights from the data and predict the unknown and make more accurate decisions. That really excites me.

Actually in my opinion, the credit loss forecasting and targeting share a lot of similarities. Although the goal of risk forecasting is helping the bank detect the risks and save some money, while marketing and targeting are helping the bank make more money.

But both of them require you to **know your customer** well.  For the loss forecast, we look at customer's FICO/Location/Product type, and use these customer information to predict the probability that this customer will default in next 2 years. And for targeting, you are looking for potential customers, so you want to know the probability of a person that might be a potential customer. Both of the jobs require you to have a solid understanding of the dealing with large amount of data, and be able to use quantitative techniques to get insights from the data to help make decision.

Therefore, How to identify the new customer and new opportunities is quite interesting to me. Marketing is a field that I am currently very interested in. Many of the skills I used in my current job will be really helpful if I can join your team.

**Consulting:**

I’ve learned a great deal during my current job in both technical skills like programming, building the models, and also communication skills. It has been a great experience.

My interest has always been the application of advanced data science techniques in the industry. But I don't want to just focus on one limited fields like the credit risk modeling in banking which is I am currently doing. So I am eager to **try some other fields** in the industry and see what I can achieve using the data analytics techniques that I have.

**Consulting** is a field that I am currently very interested in. I will have the opportunities to work on all kinds of different projects. Especially in BCG GAMMA, you have many global top companies as clients in different industry, if I can join the team, I will be using most state-of-the-art data science techniques to get insights from the data and help the clients to maximize their performance by leveraging the advanced analytics. This job really excites me.

* + Explain ML(random forest) to non-technical audience :

When I explain some of the complex algorithms to the senior non-technical audience, I will try not to go to too detailed and technical information unless they are particularly interested in, which is very rare situation.

I will explain the basic idea of ML like this:

Using machine learning is that we are trying to teach machines to “Learn from Experience”.

Machine learning algorithms use computational methods to “learn” information directly from data without relying on a predetermined equation as a model. Like regression, you have a predetermined equation like y = ax + b; For Machine learning algorithm you don't need an equation like this.

How it work is that we provide many many examples to the computer, so that they learn from their mistakes on those examples. The algorithms can improve their performance as the number of samples available for learning increases. So the ML-Algorithms find natural patterns within the data, get insights from it and predict the unknown for better and more accurate decisions compared to some of the conventional methods.

I will also educate them with some basic metrics that you can tell check to see the model results and performance; (see below: the challenges working with business.)

* + Experience working with business/IT

Since I worked on both commercial (Middle Market/Asset Based Lending) and consumer (Business Banking) portfolios, I have a lot of chances to work with business. I worked with business to discuss the data and model methodologies and results. So I think I have solid knowledge about products of our bank. I also work closely with T&O, our IT team, to discuss and validate the data.

Talk technical with business; whole cycle educate them (logistic regression: coefficients)

Talk business with IT; educate them (preprocess the financial statements; calculate the ratios)

Working with **business**:

Because the users of our model are partners of different business lines, they use our model to forecast credit loss for **strategy planning** and also **stress testing** for the bank, so I have a lot of chances to work with business partners, for the **full cycle** of the model building process.

* + Firstly, we will discuss about **business requirement**. Make sure we fully understand the business **background** (for streamlined: need a new loan-level model to replace current top-of-the-house model, provides estimation at a more granular level than the existing model), purpose of the model (strategy planning and stress testing), and how they want to use the model.
  + Then, we will get some insights of **key data elements** needed for model development from them. Since they have better understanding of the business and customer, they will provide us a list of key elements they suggested we should be looking at. There are two sets of data:
    - **Customer level data**: 100+ total, 20+ focus;
      * Small Business Banking: FICO/Location/Product type (term/revolving line of credit)/Industry type
      * Mortgage: FICO/Location/Product type (fixed/floating rates)/LTV/Loan Age; For HELOC, add first/second lien
      * Commercial: financials( EBITDA/Quick Ratio/Profit margin/Sales growth)
    - **Macro-Economic** indicators: Fed require the forecast to be sensitive to the macroeconomic variable; 20+
      * Small Business Banking: Unemployment rate/Salary growth/household obligation ratio
      * Mortgage: Unemployment rate/1 year LIBOR/HPI/CPI(Consumer Confidence Index)
      * Commercial: Unemployment rate/HPI/GDP
  + After doing some preliminary analysis of data, we will propose **model methodology** to business depending on the model purpose and available data.

We tried methodology like logistic regression and decision tree model. And found that the decision tree did not result in much superior forecasts in the sense of the forecast accuracy. So the regression models were, therefore, selected as the final models because of their advantage in **transparency** and ease of **interpretability.**

We used decision tree model as **benchmark** model to compare the predict accuracy.

* + Lastly, we will share the **model results** with the business and ask for sign-off on the model methodology and results.

Challenges:

* + Explain technical terms for non-tech audience:
    - Explain the model methodology of ML: not go into the details, just explain the **basic ideas**; see above
    - Explain model results: educate them with some basic statistical results: Logistic regression as example
      * Estimated **coefficients**: larger than 0 represent higher chances of defaulting;
      * Or: **Odds ratios** of the estimated coefficients: odds(success) = success/failure; larger than 1 represent higher chances of defaulting (the event is more likely to occur as the predictor increases);

Logit(success) = log(odds(success)) = log(success/failure); logit(success) = a + bx

* + **VIF** (Variance Inflation Factor): measure the presence of **multi-collinearity** in the model; VIF < 5 means absence of strong multi-collinearity
  + **P value**: coefficient significant level; see p value.

* + Bridge the gap between business and IT

Work with IT:

* + Bridge between business and IT; aggregate different source of data into one single master database; implement the logic from business and share and teach IT
  + We works with IT department to gather the required data elements for model development.
  + Check availability and data quality:
    - FICO: compare different source; impute if missing
    - LTV (loan to value ratio: divide money borrowed by the value of the property): important but missing; inform and help IT to improve

<http://www.1point3acres.com/bbs/thread-209651-1-1.html>

Why DS/亚麻？

Make real impact

I realized I wanted to work in a company where data science is

part of its core strategy and the company has a clear direction.

Can utilize proficient data manipulation skills and comprehensive data analysis knowledge to drive business insights

enthusiasm, willingness and ability to continually learn new things

Passion for the company and data science in general

当面试官问你有没有问题想要问……

Do you have questions for me?

<https://www.1point3acres.com/bbs/forum.php?mod=viewthread&tid=346993&extra=page%3D2>

面试结尾如何提问？我一般会问对方，

最近做的什么project最exciting。除了很少数非常junior的人会说，不能告诉你，其他绝大部分时候对方都会极其高兴跟你讲。这问题好处是，你的确可以知道对方组里到底做了什么大事，如果有不match也可以问清楚。比如你想做大数据infra但是对方跟你讲统计建模，你就需要打个问号，找剩下的人问清楚了。另外一个巨大的好处是，这时候你是主动的一方，别想着被考完了松口气，这时候一般是最动脑的时候，要快速做点笔记，记下你有更多问题的地方，深挖一下问题，比如某些选择为什么这么做不那么做，如果你有些相关想法可以提出来，一般对方会特别高兴跟你说这是个很好的问题！我们的确也考虑过，但是是这么这么。。。。。然后你再抓住发挥个1-2次。 假定你的确好好思考过这个产品，那你多少都会提出有价值的问题的。比如可以问竞争者啊，类似产品啊，类似功能啊，数据收集上的难点啊，系统啊，分析啊etc。反正我每次电面都问的面试官自己高高兴兴地讲超时一阵子。

From <<https://www.1point3acres.com/bbs/thread-111681-1-1.html>>

Why use **python**?

Python is a simple scripting language that makes it **easy to interact with data**. Furthermore, Python has a **wide range of packages** that make it easy to get started and build applications, from the simplest ones to the most complex. Python is **widely used in industry**, and is becoming the de facto language for data science in industry.

We will also use the **IPython Notebook** in our videos. The IPython Notebook is a simple interactive environment for programming with Python, which makes it really **easy to share your results**. Think about it as a combination of a Python terminal and a wiki page. Thus, you can combine code, plots and text to explain what you did. (You are not required to use IPython Notebook in the assignments, and should have no problem using straight up Python if you prefer.)