

Interview 21 - PO

Interviewee

All right. Okay, perfect.

Interviewer 1

All right. Uh, so to begin with, uh, can you give us some background information? What's your current position? How much experience you have in ML or

Interviewee

in general? Sure, sure, sure. Um, so this is Interviewee from Country X. Uh, current, I'm working as a data scientist with, uh, Company based outta, uh, Southeast.

Interviewee

And, uh, this is actually my very first data job and, uh, I've been working with a company for almost three years. Um, so, uh, my company started, uh, with, uh, started as a, uh, red company, but right now it is providing services ranging from on demand services, uh, financial services, as well as, as well as e-commerce.

Interviewee

And, uh, right now I'm working with, uh, food delivery department. It is, uh, it falls under the category of on demand services. Okay, great. Thank you.

Interviewer 1

Uh, so I'll start for the first question. What are the main quality issues that you have encountered with your data model or system so far?

Interviewee

Uh, okay. Uh, okay. I can mention two issues, uh, with.

Interviewee

With our, with the data quality. Um, okay. So, uh, I, my company, I have, uh, experience developing, uh, two major, uh, uh, ML software system. So, uh, one of them is a recommender system, as you mentioned just now. Uh, the second one is a project I'm working on right now, which is to develop, uh, budget allocation and. To help our restaurant partners to better spend their money on carrying out promotion companies.

Interviewee

Yeah. Feel free to stop me if you have any questions. So, um, I will say under my department, uh, the food delivery department, uh, we are developing a bunch of models, uh, which deal with, uh, matching between user and the restaurants. Yeah. So I will say, uh, many models are quite similar to each. Okay, so, uh, the main resource of data we used, um, is our, uh, data transaction.

Interviewee

Data. Yeah. So, uh, above the important data sources we are using is a booking, booking log, which cause attributes of, uh, transactions which were completed on daily basis. Yeah. So, uh, this booking. Uh, contains, uh, hundreds of attributes so that we can extract, uh, transaction related user behavior information.

Interviewee

So, uh, I can give you an example. Uh, we are interested in using features like, uh, uh, frequency of cooking, uh, over the past number of days. Yeah. So, um, it means our, um, uh, user restaurant matching model typically use, uh, this kind of features. All right. So it means, uh, we actually rely a lot on this, uh, booking, booking log database.

Interviewee

So one data quality we encounter, it's not me, it's, uh, all my colleagues. So, um, okay. Our booking log database is getting, uh, updated, uh, every few hours. It's not getting updated on realtime. And, uh, this booking log was ingested from our realtime. Yeah. So, um, there was a time, uh, when we noticed that, uh, the booking log for the past few days were act was actually missing.

Interviewee

Yeah. And, uh, we only noticed this issue when we, um, when we observed the weird prediction values coming out from our ml also. Yeah. So, uh, this is because we don't have a system to, uh, we don't have a solution to monitor the quality of our features. So, for example, maybe certain, certain feature value actually fall under the correct range there.

Interviewee

Maybe there are supposed to be like thousands of transactions, but maybe in a few days there are only like less than 100 transactions. Then something must be, , right? Yeah. But we don't have a system to monitor this, so we only notice an issue when we see, uh, prediction value is getting bigger, right? Yeah.

Interviewee

So to deal with this, uh, what we did is, uh, we actually, uh, fall back to the perverse version of the model, which were trained, uh, let's say, uh, a vehicle. Yeah. Cause at least the model can, uh, give us some meaningful results, although it is not trained based on the most up to date data. Yeah. Okay.

Interviewer 1

I understand.

Interviewer 1

Uh, so what you're, what you're saying is that they were missing values during training time for one of your models, not at inference time. Am I

Interviewee

correct? Uh, yes. Yes, you are.

Interviewer 1

Okay. Super interesting. Thank you. And now do you have No. Okay, perfect. Thank you. Hmm. All right. Um, and do you use any of the following data collection technique?

Interviewer 1

Uh, so you already mentioned you use data generated by, by logs as so one other system. Yeah. Uh, did you ever use, um, the services of someone who fetch, manually fetch data for you or label data for you?

Interviewee

Hmm. Okay. So for most of the models we developed, we are using databases, uh, from uh, Google Query. So, uh, we mostly write queries to, uh, extract the features.

Interviewee

Yeah. And, uh, regarding the question you asked us now, yes, we do. Uh, we also have a content team who are dedicated on, like making annotations for our text based data or image based data. Yeah, but for me, uh, I have experienced training, text based or image based model, but those models didn't, uh, go into production, so I'm not gonna call them as software system.

Interviewee

Okay,

Interviewer 1

I see. Uh, but did you have any issue with the, um, data labelers? Was there any Yes. Okay.

Interviewee

Yes, yes, yes, we do. So, um, okay. Uh, As you know, since we are providing, uh, food delivery services, so, uh, this require us, uh, this re this will require, uh, our restaurant partners to upload their manuals, uh, including both, uh, text-based description as well as dish images to our platform.

Interviewee

Um, yeah. Uh, however, uh, many. The many of our restaurant partners, they are actually very small, skilled, uh, small, skilled, uh, restaurant merchants. And, uh, they might not have a good knowledge on how to run a proper business. That is why, um, the images they submitted to us and, uh, have a bad quality. That is why we were building a system to like help us automatically filter out the low quality images.

Interviewee

Alright. So you can imagine this will require our, uh, annotation team to help us, uh, flag the low quality images in terms of, uh, things like, uh, or maybe size of the image, whether the image is a crop or not. Yeah. Um, but at the. Um, our data label, our data labelers, they might have a different, uh, standards in their mind when, when they, uh, work on labeling the images.

Interviewee

Yeah. So, so even for me, um, uh, if, uh, if, uh, image has, uh, a shallow tax of fuel, it doesn't mean that the is a blurry, so maybe the user maybe, Our restaurant partner is taking a closeup shot to make the image look nicer, but, uh, that might be considered as blurry by our, uh, image labeler. Yeah. So, okay. I see.

Interviewee

Um, yeah, yeah, yeah. So when we see the images and the labels, we notice that, um, uh, the labels are not consistent. So, uh, we have to like, reach out to the annotation team a few times to, to, uh, improve the quality of the labels. Yeah.

Interviewer 1

Okay. I see. And do you use this image with the annotation to train other model or it's only for the application you're using?

Interviewee

We use the labels mainly for training, the quality, uh, quality detection model. But, uh, this is not a single model because we want to identify the quality of images, uh, under different criteria. Like, uh, what I mentioned just now blur is size, et cetera. Uh, and, uh, we also don't want to, uh, see, uh, any watermark or non-food related item in the image.

Interviewee

Yeah.

Interviewer 1

Okay. I see. I see. But in the end, your, your images, will it be used, will it be ingested by a model or it's only for. Like for the application, um, where we present the restaurant,

Interviewee

no, uh, the images, uh, will be consumed by our, uh, image classification model. Yeah. Okay. Okay, perfect. Yeah. Yeah, yeah, yeah. But, but yeah, you are, you're right, you're right.

Interviewee

Uh, after our model detect detected any, like local, uh, our restaurant, our restaurants,

Interviewee

yeah. Okay. I see. Okay. Thank you.

Interviewer 1

Um, have you ever measured the quality of your data and or tried to improve it?

Interviewee

Manager the quality of the data. Excuse me. Can you hear me? Yes, we can. Okay. Okay. Good. Good. Yes. Uh, yes, yes I did. But uh, I'm gonna say, uh, it was very limited. So for example, for the images, it's ministry through eyeballing, eyeballing through the images to check the quality of the.

Interviewee

Yeah. And, uh, regarding the numeric, uh, features, uh, what I have done is to, um, you know, compute the very basic statistics of the, uh, numeric features to, um, get the things like a a mean value standard deviation range, what's a mean at the max? And, uh, make sure the, the. Data is in the correct range as expected.

Interviewee

Yeah. But, uh, this is only like a one time exercise. I never this, uh,

Interviewee

okay,

Interviewer 1

perfect. Thank you. Uh, is there any other data quality should be missed that you consider relevant?

Interviewee

Uh, um, yes. Yes. Uh, there is another one. So, um, The issue is still with, uh, booking log I mentioned just now. So, uh, yeah, our booking log, uh, we will, uh, record the user id, restaurant ID as well as a brand id, uh, with, with respect to, uh, any, uh, food delivery transaction.

Interviewee

So, uh, so what's the difference between, uh, restaurant ID and the brand id? So restaurant is more like an. So, for example, mag McDonald. McDonald can have many hours in Country X. Yeah. But all these hours should fall under the unique brand. So, so you can imagine that, you know, a data record, when we see both brand ID and the restaurant id, we expect the two IDs to be consistent.

Interviewee

So it doesn't make sense for us to see a KFC restaurant ID under, uh, McDonald brand. Yeah. But uh, in reality we did observe that there is this kind of inconsistent record in our database and, uh, it's quite troublesome for us to tell which one is reliable. Shall we refer to brand id, shall we refer to restaurant id?

Interviewee

It's very hard to tell unless we, uh, dive into the, uh, dish items. Uh, with, uh, booking. All right. So, uh, what we actually did is, um, we investigated, uh, how much percentage of such inconsistency is there with our database, and we saw the, the percentage is pretty low, 3%, 2%. And uh, we decided to just leave, visit.

Interviewee

Yeah. Okay. How this is other data quality issue. Yeah.

Interviewer 1

Okay. I see. Thank you. How did you detect these inconsistency? Maybe you mentioned

Interviewee

it, but I didn't. Yeah, yeah, yeah. Cause uh, we have another database which record the mapping between the restaurant and the brands, right? So based on this mapping database, we can check, uh, out of all the booking data, which record has consistent brand and, uh, restaurant pairs, which one has inconsistent?

Interviewee

Brand and the restaurant pairs. Yeah, that's Ed. Okay.

Interviewer 1

And in the end, did you drop the, uh, corrupted, um, instance, or you

Interviewee

fixed them? Uh, no. Uh, I, I didn't, I didn't fix them cause it's gonna take too much effort. Yeah. So we decided to just fix it. Yeah. Yeah. Okay. So, so you kept it basical. Yeah, yeah, yeah, yeah, yeah.

Interviewee

I, yeah, I kept it, uh, since the percentage is pretty low. Yeah. Okay. Thank you.

Interviewer 1

Uh, how do you evaluate the quality of models? And as a reminder, quality is not only defined by ML performance, but you can also consider, uh, explainability robustness, um, scalability.

Interviewee

Okay. Okay. Okay. Um, okay. I will say in my firm, uh, our ML models are used for making micro microdecisions microdecisions refer to, uh, which restaurants are supposed to be pushed to, to, uh, to a particular user.

Interviewee

All right. So, uh, and we are building, uh, models using. Sorry, I, I might need to switch to another, uh, earphone. Uh, just gimme a sec.

Interviewee

Can you hear me? Yes.

Interviewee

Sorry. Can you hear me? Yes,

Interviewee

can you hear us? Cause, cause I cannot hear you. But we can, uh,

Interviewee

sorry, can I speak something? Yes. No,

Interviewer 1

I'm talking

Interviewee

right now.

Interviewer 1

Can you hear us now?

Interviewee

Uh, can you speak something? Yeah, can you hear us? Oh, okay. Okay, good, good, good. Okay. Okay. Okay.

Uh, sorry about that. No worries. Okay. Yeah. So, uh, as I mentioned just now, uh, our software, our ML models are mostly used for making macro decisions and, uh, we use models. Such as, uh, uh, s and new networks. So imagine these algorithm, they have a poor, uh, explainability.

Interviewee

And, uh, to be honest, uh, our stakeholders mainly care about the business. They don't really ask us for explain the, the logic of the. Yeah, . So they're okay. Uh, they're okay with us to, uh, treat the machine learning as a black box. So, uh, I understand that this can be quite different from the machine learning solutions built by, uh, consulting firms.

Interviewee

Alright. Yeah. Um, so what else? Okay. How to measure the quality of, uh, missionary model. Um, okay, so, So, apart from, uh, explainability, uh, we mostly, uh, measure our machine learning, uh, the, the quality of our machine learning solutions based on latency as well as uh, accuracy. Alright. Um, so in terms of latency, um, okay.

Interviewee

Our models can be categorized into a live model and the batch model. So live model means, uh, we have to like, make prediction on the flight whenever the user log into our app and uh, he or she opens the recommendation page, we need to like, uh, give them the recommendation. We send, uh, I don't know, 2020 millisecond.

Interviewee

Yeah, something like that. Uh, okay. Yeah. So of course for this kinda, uh, models, we have to like guarantee the guarantee the, the, the, the, the, the latency. Yeah. So, uh, this requires that, um, we cannot use too many features. And, uh, also, um, if you use techniques like, uh, feature. Uh, we cannot, uh, make the processes of, uh, feature pre-processing through complex, uh, when we serve the model in production.

Interviewee

Uh, yeah. So this is regarding latency. Uh, however, for batch model, uh, since, uh, there is, uh, very low latency requirement. So, uh, as long as we can, uh, produce our. Uh, before the deadline, uh, it is. Okay. Yeah. So this is regarding latency and, uh, lastly, uh, the is is probably the most important, uh, criteria for measuring the quality.

Interviewee

Uh, so we, we'll look at, uh, accuracy, uh, Yeah. But of course, different models, they'll use different metrics. So for classification, we might be using R C C or, uh, . Exactly. Yeah.

Interviewer 1

Okay. Super interesting. Thank you. Uh, so just to summarize, you have for latency, you have, uh, online models and offline models. And for the online one, uh, to, in order to make them faster, you, you have less feature, uh, basical.

Interviewee

Yeah, use, use fewer features and, uh, make the process of, uh, per feature processing simpler. Okay.

Interviewer 1

See, thank you. Super interesting. Yeah. Um, have you used existing bench benchmark models for quality aspects to evaluate your model? So did you ever compare your models to a baseline to evaluate its quality?

Interviewee

Yes. Yes, yes, yes, we do. .

Interviewer 1

Okay. Uh, which, which benchmark do you

Interviewee

use? Mm,

Interviewee

let me see.

Interviewee

Okay. So, uh, lemme give you an example, uh, based on the current project, uh, which I'm working on. So, uh, just to recap, uh, this project is for, um, Helping our restaurant partners allocate their budget, uh, in a smarter way. Alright, so in this case, our baseline is actually, uh, random allocating the budget. So imagine you have a 10 K to spend over the next two weeks, right?

Interviewee

And, uh, uh, you have a choice of 20% discount or 30%. Right. So, uh, in this way our baseline will be, uh, we can sample, uh, maybe 100 K users and, uh, we randomly split them into 50 50 and then we give one group 20% discount, another group 30% discount. Yeah. And then, uh, at the end of the campaign, the measure.

Interviewee

How much revenue we generated from the campaign, how much spending is, how much spending was incurred during the campaign. So we take this as, as our baseline. Yeah. Okay. I see. Thank you.

Interviewer 1

Um, have you ever assessed a quality of, uh, ML system with the user of the system?

Interviewee

With the user of the system? You mean?

Interviewee

Uh, asking the user, uh, if they. Getting what they expect to get. Yes. Right. Uh, no . Okay. Yeah.

Interviewer 1

Okay. Thank you. Um, is there any other quality issues during the evaluation of your model that you'd like to talk

about?

Interviewee

Um, okay. Um, we have been facing, uh, contamination. Uh, then we evaluate, uh, the current budget allocation solution we are building.

Interviewee

So, um, uh, yeah, because, because, uh, the idea of the solution is to identify the users with a higher sensitivity to promotion. So just imagine you have a limited amount of budget and, uh, it's not enough. Cover all of the users. So you have to select, all right, so in this case, uh, you don't want target the users who will transact any way or without the discount, right?

Interviewee

So you want to identify the users with a better potential of conversion if you gave them a discount, right? So in this case, when build model, we will want to accurately measure the. Of our promotion with respect to users, we, users who are exposed to our desktop. However, the problem with our approach is that in our company there are many campaigns, uh, uh, which can be ongoing in parallel.

Interviewee

And, uh, some of our campaigns are not managed by our team. So we don't have the right to interpret, we cannot tell them. We are carrying out some AB tests. Don't run your campaign . That's not possible. Yeah. So, uh, the only thing we can do is that we just assume that the campaigns managed by the other teams, they are constant, they are, uh, exposed to all the users.

Interviewee

Yeah. I see. Yeah. But you can imagine the impact we measured for our own campaign might not be accurate because these users, they have been exposed to some other promotion. Yeah. I see. That's

Interviewer 1

a really good point. Interesting. Thank you. Um, what are the challenges you have encountered when deploying machine learning software?

Interviewee

Deploy the machine learning system. Hmm. Okay. So, uh, honestly, I do not have much experience on deploying a machine learning system in my firm. Uh, so, uh, in my company we build a internal system which help. Our data scientist to, uh, deploy our system. So, uh, we, uh, by using the system, uh, we have the freedom to deploy user model, our batch model.

Interviewee

Yeah. So I do say one challenges, uh, data scientists, uh, usually face is to, um, ensure the maintenance

Interviewee

is to latency. Yeah.

Interviewer 1

Yes. I, I missed your last sentence because my, my zoom. Yes. Thank you. Sure,

Interviewee

sure, sure, sure, sure. Uh, okay, I'm gonna save One major challenge, uh, faced by the data scientist is to ensure a good latency when they, uh, model. Okay? Yeah. It's quite common that after they deploy the model, They measures the latency, it's something like, uh, over 100 millisecond.

Interviewee

So the require, the requirement could be 50 milliseconds, so it's far from the requirement. So they'll have to, you know, figure out how to simplify their, uh, data processing pipeline, how to simplify their feature set. Yeah. Yeah. Okay. I see. Thank you. Yeah.

Interviewer 1

Um, do you have any, do you have experience in the mainten.

Interviewer 1

Of machine learning software system

Interviewee

maintenance. Um, can you clarify the definition of maintenance? Yes.

Interviewer 1

Uh, so when you deploy a machine learning software system Yeah. What's it deploy? Uh, then you are in the maintenance phase of your machine learning software

Interviewee

system. Yeah. Okay. Okay. Yeah. So, uh, after we deploy our learning model, we'll, you know, keep monitoring the performance, uh, performance of our model.

Interviewee

Uh, so this performance refers to business. So case can be conversion rate or, uh, improvement on the cost efficiency. Yeah. So, um, If, if, um, we, if we see constant, uh, if we see consistent performance, uh, we'll not wanna, um, you know, make large change to our, to our models. And if we see some variation of the, of the

model performance, which can be weird, in that case we'll investigate your troubleshoot, uh, what goes wrong with the.

Interviewee

Yeah, in many cases it can be, um, it can be due to external factors, like someone runs some, uh, very LA companies or it can be, uh, data issue like, uh, what I mentioned just now. Okay. Okay. Thank you.

Interviewer 1

And, uh, so that's, that, that's what's your answer is really close to my first question. Um, how do you ensure that the quality of a machine learning software system does not decrease over.

Interviewee

Quality insurance system. Uh, okay. The main idea I have is through, uh, comprehensive, uh, monitoring system.

Interviewee

Yeah. Okay. So yeah, so this monitoring system should cover both, uh, data quality and the business metrics. So, uh, I will say for now, uh, at my team, uh, we closely monitoring the business not for.

Interviewee

Okay, perfect. I see. Thanks.

Interviewer 1

Uh, did you add issue with one of the following quality aspect? So, fairness, robustness, explainability, scalability, privacy, and. So one of them?

Interviewee

Yeah.

Interviewee

Fair?

Interviewee

Yes. Uh, okay. I have one point. Uh, in regards to fairness. Um, this is a issue we notice in recent period, however, we haven't, you know, come up with a good. Um, so, so, uh, our team noticed that there are users who have been consistently predicted allocation of promotion. So, so someone might, might, might be this, uh, lucky person who, okay, every time this guy is receiving promotion from us, cause model predicted that, oh, he's, uh, yeah, yeah.

Interviewee

So our business team, uh, uh, actually had this response, had this feedback. So this is not fair. For a lot of other users. So is there a, is there a way for us to deal with it? So it also doesn't make sense for us to say, Hey, this user has been receiving promotion from us for past month, months. How about we just exclude them?

Interviewee

That, that, that doesn't, that doesn't make sense. So yeah, we haven't figured out a solution. Okay.

Interviewer 1

I see. And, uh, these, this group of user received a lot of promotion. Did, did they have some share feature? Like, uh, was it all males or all? I dunno. Okay.

Interviewee

Uh, cause we've been using really quite a lot of features, so it's very hard for us to summarize a pattern for certain users.

Interviewee

So based on my observation, it feels that if you are very at. You'll keep on transacting. You will not, you will not be targeted by the model if you never transact. No matter if we do discount to you or not, you'll also, you also won't be targeted. So it's a middle class users who typically get targeted by the model that, that's a pattern I observe, but it might not be accurate.

Interviewer 1

Okay. Thank you. Uh, did you have any issue with the privacy of your data or not? Privacy?

Interviewee

No. No. Okay. Perfect. Yeah. Yeah. Cause all the, all the data we use, uh, comes from our internal database and, uh, it is collected by our applications. Yeah. Okay. Thank you.

Interviewer 1

Um, so I have like one or two last question. Uh, so in your opinion, what is the most pressing quality issue researchers should try to solve?

Interviewee

Uh, Okay. Uh, I will say data drift. Wait, wait, sorry. Uh, I have another point. So, um, the, the, the most important issue, uh, we are dealing with is, uh, class imbalance. Yeah. So, uh, since we are working with a lot of matching problem between user and the restaurant, and you, you can imagine one user. Just place a few order made, made a few booking with a, with a single restaurant.

Interviewee

So that is why in our data set, it is very common that the possible example is only like below 5%. Yeah. So, okay, I see. Yeah, yeah, yeah. So in that case, we are actually dealing with a data service, a very high capacity. So in this. Um, most of our features, they only give us a, a very, uh, they can only give us big signals.

Interviewee

Yeah. So in this case, how to, uh, improve the quality of the model performance. Yeah. So that is something, uh, I'm interested And, uh, another, another topic I mentioned just now is regarding data. How to identify the shift of user behavior, um, due to maybe, uh, seasonality, uh, or maybe, uh, macroeconomic, uh, situation.

Interviewee

Yeah. These two items, uh, you'll say, uh, they are most critical, uh, for the cases of, uh, real world applications. Yeah. Okay.

Interviewer 1

Um, Do you have any, do you have any tool to detect data drift or seasonality in your data set? Mm,

Interviewee

no. I don't have, it's maybe through like a simple cooperation for the, for the statistics and then maybe we make some comparison if there is any difference.

Interviewee

Okay. If there is, we will try to. But if it is something we cannot explain, then we probably just leave it . Yeah.

Yeah. I see. Okay. Thank

Interviewer 1

you. And, um, do you have any other comment about the quality of ML system,

Interviewee

quality of the ML system?

Interviewer 1

Anything else we, you didn't mention and you'd like to, to mention It basically,

Interviewee

Okay. Um, I'm gonna say the cost efficiency of our MO system is not very transparent to me. Maybe it is transparent to my manager, but not to me . Yeah. So, uh, I'm not fully aware of how much money we are spending, uh, by, uh, executing one. Yeah. And maybe, uh, retraining the model also cost us a lot of money, but, uh, we are not aware of that.

Interviewee

So yeah. Is it worthwhile for us to develop a complicated system? Uh, could we even break even? Yeah. So this is something I'd like to know. If there is some simple tool which can help us with this, then it'll be very helpful. Okay. I see.

Interviewer 1

Uh, so with the es esq query, the issue is you are fetching a lot of data, which may cost a lot of computing power, something like that.

Interviewer 1

Yeah.

Interviewee

Okay. I see. Cost. Cost. We are querying data using the Google services and, uh, it's never free . Yeah. And our query can like, uh, can be processing hundreds of gigabytes of data. very common. Yeah.

Interviewer 1

Okay. Okay. See, so you would like some solution to, to, to make it clear how much it'll cost you, more or less.

Interviewee

Yes. Yes, yes, yes.

Interviewer 1

Yeah. All right. So that, that's all. Thank you all. Thank you for all your answer. So it's really a pleasure to have you, and I'm sure that everything we discussed will be, uh, really useful for our, our

Interviewee

study. Uh, yeah. Uh, yeah, that will be great. Yeah. Yeah. Uh, please, uh, after you finish your study, uh, do share, share with me your report.

Interviewee

Sure. Absolutely. Yeah.

Interviewer 1

All right.

Interviewee

Thank you. Have a good night. Thank you. Thank you. Bye. Have a nice. My.

Created with the Delve Qualitative Analysis Tool (<http://www.delvetool.com>)