Interview 5 - Rached

Interviewer

Catalog of quality issues and machine learnings, machine learnings of our systems. Uh, so what is a quality

issue, basically? Any issue you have encountered while building ML system that affected the is quality and

what is a ML system? Um, Any system that has ML component. So you could think, for example, or

recommend their system for net Netflix, it has a ML ML model in the center that does prediction and the

recommender in itself is an ML system.

Interviewer

So when there's a ML model in it, you have an ML system can sure. Think of it like this

Interviewee

it's straightforward. Yeah.

Interviewer

All right. And why? Because we want to, uh, guide future work on improving the quality system. Mm-hmm um,

so I will ask you some question to set up the interview. Um, so we have your permission to record the

interview.

Interviewer

That's

Interviewer

great. Yes.

Interviewer

And I will ask some background information from you, uh, just to get to know you better. So mm-hmm could

you tell me, what is your current push position? Uh, how much experience you have in general and specific to

machine learning and these kind of

Interviewee

information? Sure. Um, so I'm a data scientist at Company Y.

Interviewee

Um, I officially started about two months ago, so I am pretty new there. Um, before that I. Did my master's of

management and analytics, um, at University X and, uh, while there I was working, um, to develop like, Next

to my study, I was, uh, still working with company X, um, doing some machine learning stuff for about like six

months, 6 to 8 months.

Interviewee

Um, but so that's the most of my machine learning experience. But before that I was working mostly in like

data analytics, um, in research context, so, yeah.

Interviewer

Okay. Okay, great. Thank you. And so the next question is for what purpose do you use AI at your company?

Move? I know move AI, but

Interviewee

yeah. For what purpose do I, what for,

Interviewer

do you use artificial intelligence at your company at move?

Interviewee

Okay. At move. Okay. So, um, right now we are building a recommender system actually, uh, for a I'm actually

not sure if I'm allowed to, to say. The the client's name. I'm not sure what the NDA terms are with you.

Interviewer

We will anonymize the transcript. So you don't have to worry, but, but yeah. Yeah, you, you are not forced to

refer to a specific customers.

Interviewer

Mm-hmm if you do, we will

Interviewee

remove them. Okay. Okay. For sure. Yeah. So if, if you want, if you want to not mention them, go

Interviewer

ahead. But if suddenly you

Interviewee

mention them, we will ignore them. Okay. Cool. Okay. Cause I'm okay. I'm holding myself back a little bit. Um,

but yeah, so we're working actually on developing a recommender system, um, to recommend products, um,

to users.

Interviewee

Um, that's essentially what we do, but we're using, um, a, we're not building a, a model from scratch, but we

using a, a solution that already exists, which is the Amazon personalized. Um, And yeah, that's mostly what

I'm working on. Um, and that's where I use AI.

Interviewer

Okay, perfect. Thank you. Mm-hmm uh, could you do a previous description of the type of data you're using?

Interviewer

How do you collect this data and for what problem do you yeah. Use this data?

Interviewee

Yeah, for sure. So actually working with, um, mostly. Web interactions data. So, um, this is literally anything

that says, um, what did this user browse at? What time? So what page ? So we have lines of that, um, that we

collected from the, the client already had, um, the data available because you were working with a data

provider for that.

Interviewee

So that was in. Too much of an issue collecting that. Um, and then on the other hand, because we are recommending, uh, recipes, right. Uh, we are also collecting and gathering data on, um, users and like demographic information as well as like, um, preferences in terms of what, like what they like to eat, or also some restrictions because therees, or, um, cultural things, um, And we're also collecting data on the taxonomy of those recipes.

Interviewee

So like for each recipe, what are the ingredients on there and how can you categorize them? Okay. So far.

Interviewer

Thank you. Uh, so if you don't mind, we will jump to the body of the interview, like the main content. Yeah. Perfect. Uh, so. I think I said it in the emails, I will go through each phase in a machine learning workflow.

Interviewer

Mm-hmm so data collection etc until, uh, you deploy your model and you monitor it. Yeah. And I will ask question regarding quality issues you may have encountered in each of these phases. Okay, perfect. So we'll start with the general question. Um, Sorry for the delay. So what are the main quality issues you have encountered with your data or model, um, or system, even ML system so

Interviewee

far, so far quality issues, um, in the development of the model?

Interviewee

Um, we, I I'm, I'm going to talk about quality, right? And I'm gonna say that right now. I think about it. Like the performance of the model or what other, what are the definitions are you using when you say quality?

Interviewer

All right. Um, so you, you can have quality quality of the model. That's that's for sure. One definition.

Interviewer

You can also think of like explainability, robustness, um, scalability. What else? I, the short list.

Interviewee

Yeah. Okay. Go for it. First thing that jumps to mind then is definitely the explainability, because we're basically using a black box tool in a way, um, personalized. They designed it to be something that anyone who doesn't know machine learning can use.

Interviewee

Um, but when you're actually using it, you realize that when you want to understand the behavior of their model, you don't have much information. They just tell you that this is, um, a neural network, a hierarchical neural network, but that's what the information that, you know, uh, you also don't know what exactly you can't really map out the weights of the features that you have put into the model.

Interviewee

You don't know exactly what's contributed to what? So while we are developing this model for the client and we wanna get them, give them results, we are also facing the issue. Whenever we want to put something out there. We want to understand the impact because we can't, we can't just tell the client here it is.

Interviewee

And then we have questions and we don't know, uh, how to explain to them what changed when we added X feature or when we removed X feature. So we have to have basically like a test bed of experiment, um, where we have to, you wouldn't necessarily do this. If you knew, if you could expand your model with like feature importance and things like that.

Interviewee

But for us, it's like, uh, what if I add one more column of data about the user? What is that gonna change in the recommendations? And so you have to go and design a whole experiment every single time you wanna test something. Um, so, so far we've worked on it for like two sprints and we've already had to do three experi.

Interviewee

Which don't seem like they bring value directly to the client. It's more for us to understand the client doesn't see that. So I think that's like a, an issue of explainability. Um, I guess I'm also thinking about, um, when, when you developing a machine learning systems, when you wanna kind of like, I guess see the impact of.

Interviewee

You're working a lot with the user experience and the user interaction of your model. And right now, for us, that's not always set up from the get go. We don't really know what, how exactly it's going to be used or in what context, how many recommendations should we give? So that's been a little bit of a, I'm thinking it's a quality issue in terms of the fact that we have to make a lot of hypothe processes that might change throughout the whole, um, development.

Interviewee

So. Yeah. Okay.

Interviewer

Um, I'll ask you some questions just to know, to see if I understand correctly. Yeah, for sure. So you mentioned you had explainability issue mm-hmm and you were using per personalized mm-hmm and it's a service that automatically creates a model for you and you think it's a neural network, something like that.

Interviewee

Uh, we know it's a neural network. Yeah, because that's what they said. Mm-hmm um, But again, it's a black box. Okay. Yeah. Yeah. Like we don't, we don't even know like, okay, how many hidden layers we think it has. Do they, do they change that for each client? Or is it the same for everybody? Um, what is their strategy when a new user comes in and we don't have any data on them, do they just give the most popular ones?

Interviewee

Is that gonna change or is it always the case? It's more like those kind of questions.

Interviewer

Okay. I understand. And, and the issue, the explainability issue for you is that you don't know how to improve your performance because you don't know the architecture. You don't know how,

Interviewee

right. Okay. Yeah. There's that? I think there's definitely that side, but there's also this side of like, um, usually say you have.

Interviewee

A very like simple linear regression. And you can kind of look into the coefficients to kind of explain what played into what, but in this case we can't tell the client that if you would like this item to be recommended more often, this is what you should do to it. We don't have the ability to like drive business insights after that in a way.

Interviewee

OK. Okay. Yeah. So it's like a two-sided. Okay.

Interviewer

I see. I see. I see. Great, interesting. Uh, perfect. So we'll move on to the question regarding data collection. Mm-hmm yeah, it's a, it's a really general question. How do you collect data to train your models?

Interviewee

How do we collect data to train our model? Um, they don't look it in our case.

Interviewee

It's uh, Do you mean like the architecture of the collection or like from where we're getting it from?

Interviewer

Exactly where, where it's. Yeah.

Interviewee

So at the moment, actually, we had a lot of, uh, issues accessing the data, but it wasn't really, I wanna say you could say quality, I guess you could look at in the long term, is that the data provider that, um, the client is working with, who is supposed to give us the interactions for the users?

Interviewee

Um, Actually did not. They told us that they had three years of data. And then, uh, because of some miscommunication, we realized that they actually did not have three years. They only had two months of data. Um, so, um, when we were collecting data, a lot of it was done through like, and, and their API was broken.

Interviewee

Um, because I don't know for whatever reason we had to go through a lot of like, okay, we'll give you the extract. August 9th. And then they give us a PSD file via email, or like, okay, now we'll give you the extracts for August 10th and so on. And so on. That's how we've been collecting the data until last week when we finally got access to the, but other than that, because we're still working in the sandbox environment and not quite of decline.

Interviewee

Um, we have like a dump of their database. Um, In, um, like in AWS, if that's where we get stuff from as well. Okay. Perfect. Thank you. So we're not, we're not actively like collecting data. We're just receiving what's already out there. Yes.

Interviewer

Perfect. I understand. And did you have any quality issues with, with the data you receive?

Interviewee

Uh, yeah. The main thing being that, um, The web interactions, as well as the, like the, yeah, I'm just gonna say names here, but basically we did a provider was, uh, early at, and then they have Sy as their data based thing. And the information between the two is actually not integrated, so required a lot of work for us to map.

Interviewee

Um, earlier I, the symphony ID. Um, because we needed to identify, right. Like if I wanted information about a user, um, I needed to be able to map that to a particular idea. So there was a lot of, uh, um, we were for the longest time before it was integrated, we were working on like mocked users. Um, So now that we're adding the user, we're actually trying to add them this week.

Interviewee

And now that we're doing that, we, we we're a bit uncertain about what the, the outputs are going to be if it's going to match, or if it's going to be like very different from the kinds of predictions that we've had in the past. Yeah.

Interviewer

Okay. Okay. So I think there two problem. One is you have user data in different database and it's difficult to match one profile with another one from different database mm-hmm

Interviewer

And another problem is, uh, you have, well, you have mocked users, right? And, uh, you are, I'm not sure if I understood this part correctly with the, the mocked users. Oh,

Interviewee

matching

Interviewer

users. Okay. Maybe I just simply misunderstood, right? Yeah.

Interviewee

I mean also like the issue with that was that when you, when you try to map match users, um, we have the, the, the early again ID is going to be different for someone who, uh, connected on their phone and then they move on to their computer.

Interviewee

And now we've. Something like, uh, that's where most of the, the mapping issues come from. Okay.

Interviewer

I understand. Perfect. Thank you. All right. Um, so moving on to data preparation, um, data in data preparation, I, I consider data cleaning and data transformation, so, okay. Data transformation, like normalizing these kind of stuff.

Interviewer

All right. So which tools and framework do you use to prepare your training

Interviewee

data? Um, actually, um, all of the, all of the data preparation we've been doing it using, um, like we just use as a, to make sure that we have a pipeline of transformation that was following like data engineering, conventions. What's the name?

Interviewee

Sorry, Kendra. Um, I can write it in the chat.

Interviewee

Okay, thanks. So that's the main tool that we've been using because it was, it just allowed us to, um, take sales from a notebook that was transforming data and automatically pushed them to become nodes in a pipeline, which was really useful because whenever we are making changes to something or we're getting new data, we just have to rerun the pipeline again.

Interviewee

And then it gives. The, the primary data set that we want, uh, without having to like me too much the things.

Interviewer

Okay. Thanks. Um, I will move on to the next question. Um, yeah, this one, um, what are the pain points that you repetitively encounter when you are preparing data for machine learning?

Interviewee

Uh, pain points that repetitively, um,

Interviewee

I. It sounds super silly or it sounds like it's nothing, but like just dealing with, um, matching the format of, again, the two different data, um, like data, data sources and identifying. Tweet particular layer. Are we changing names? And what if we do this? What are the implications, if are changing the types of columns and then change them back?

Interviewee

At what point do we do it? When should we not do it? It's a lot of like, back and forth on that. Um, but yeah.

Interviewer

Okay. I understand. So data integration is, is a big issue. I think that,

Interviewee

yeah, understand. Yeah. Mm-hmm and is there. Actually also like the, the documentation that we get from the, the fact that we are getting the data from third parties, but not actually like in a way collecting it ourselves or generating ourselves is a little bit like, um, silly mistakes can happen when you don't understand what ID customer means versus user ID or things like that.

Interviewer

Yeah. I heard you yesterday.

Interviewee

Oh, you. It's annoying. Yeah.

Interviewer

Right. Uh, is there any other data quality issue you missed and you consider relevant?

Interviewee

Um,

Interviewee

no, I can't think of part that's perfect. scarcity of the data and, uh, yeah.

Interviewer

Good. Thank you. All right. And moving on to model evaluation, um, how do you evaluate the quality of your models?

Interviewee

that's the question that we're all asking ourselves at the moment. Um, how do so actually we're because it's the recommendation system that it's eventually going to go like online right now, everything that we're measuring is actually off.

Interviewee

Uh, measurements. So like send metrics. Sorry. So we're, instead of, we're not looking at like group people interacting with it yet or anything, it's just like we're looking at coverage. Um, and basically how much of the, uh, recipes in the whole database could get recommended in a, by the model. We're trying to get that as high as possible.

Interviewee

So we know that we have a variety of recipes and not just like the most populars are coming out. Um, and also we're trying to evaluate it based on, um, the relevance, like the precision, like after how many, like how relevant are the, the first 10 items, for example. And that's, again, another issue is that those are actually metrics that Amazon personalized can.

Interviewee

But we don't have a lot of insight into how they calculate those scores or we, we kind of acting with like our best judgment, like, oh, how do you usually calculate coverage, maybe like this and hoping to imagine that. Um, but, uh, we are actually developing like a, a small, like testing with like real users.

Interviewee

Um, we kinda call it like the internal pilot. Which is basically, we're trying to evaluate our model from, by getting feedback from real users. So we're basically pushing some re giving some recommendations to people once a week asking them, um, how they heal about those recommendations. And then that's how we're trying to gauge like the quality of it.

Interviewee

Like that just actually makes sense. We're not only looking at things like, um, how, like how many, uh, how many items are recommended, but also looking. Would someone actually click on it, which I think that we can't actually measure unless we go online. Um, yeah. So we're developing that, uh, testing thing, which is

also taking a little bit of like extra time to just set up, plan the experiment again, and, uh, set up all the variables, set up the, the user, the UI and all of.

Interviewer

Okay, thanks. Perfect. You, you already answered a question. I was coming to ask, so it was regarding a user acceptance,

Interviewee

so perfect. Yeah. That's uh, because I don't know, I feel like in this case, because it's, so when you're talking about recommendation, it's you do need that kind of like, uh, qualitative feel, right?

Interviewee

Like. You're not just looking at yes, no. And yes, no, but you wanna know you, you want a little bit more human input into the development of it.

Interviewer

Yeah. It, it, I guess it, it can be difficult to evaluate their ranking engine I'm. Yeah.

Interviewee

Yeah.

Interviewer

Um, okay. So are your models used in scenario scenarios involving different groups of.

Interviewer

For example, people of, of different skin color with different sex. And it's a weird, good. Yeah. Anyway,

Interviewee

um, how are, how is the model used by different group of people? You mean like, how are the outputs of the model used or like how, like, if I am using it versus the other data scientist is using it, how no, no

Interviewer

differently.

Interviewer

Or I meant, I mean, yeah. Your your model and outputs prediction for different groups of people.

Interviewee

Okay. Um, so actually right now we're only planning to build one model because we know that we have a target, um, audience that is pretty homogenous. Um, in that sense, everything from the get go was, uh, built from the idea that this model should be able to cover.

Interviewee

Um, The majority of these people, because they all have the same profile. But I guess if we were to think in the context of this, if we wanted to, um, if he was geared for a more diverse group of people, we probably would've identified like user segments and, um, identify if we needed to build different models, completely different models or.

Interviewee

Yeah. Or dis parameters yeah. In this case? No. . Yeah.

Interviewer

Perfect. Thanks. Uh, have you encount encountered any other quality issue during the evaluation of your models?

Interviewee

We haven't fully gotten to that point yet. So in this particular project, um, I can't say much.

Interviewer

Okay, perfect. So the next phases are regarding, uh, deployment maintenance.

Interviewer

Yeah. These two phases. Maybe you, I've not lot experience there, but if you, I mean, I would ask you question and if you don't know the answer, uh, we will move on to the next one

Interviewee

. Yeah. Yeah. Okay.

Interviewer

Um, so do you monitor your ML models once they are deployed?

Interviewee

Um, can't answer.

Interviewer

Okay, perfect. Um, do you have some models that have gone stale after some time?

Interviewee

Interviewee

Can't answer for this either. Right. But, uh, I'm sure. Yes. That is something that happens.

Interviewer

Yeah. Um, have you ever faced problem related to this? Oh, sorry. Skip apart. Okay. Uh,

Interviewer

Sorry for the, a bit of trouble. Um, did you have ever add a prob a model that perform well locally, but poorly

once the deploy?

Interviewee

Uh, I am anticipating something like that with the recommendation engine that we just test offline for now, but,

uh, once we get the answers in a couple of weeks, well, no,

Interviewer

OK.

Interviewer

Perfect. Um, Yes, I think that's it for the, uh, model deployment and maintenance question. Uh, alright. I have a few question outside of the scope of the workflow of, of ML workflow. Mm-hmm um, have you ever faced problem related to the scalability of three model,

Interviewee
um, scalability

Interviewee
for this particular project? I feel like we haven't really gotten to a point where we've had to like put it to for like a huge amount of users or where we have too much data that we're trying to predict from. So, no,

Interviewer
thank you. Um, is robustness a significant quality issue when building ML.

Interviewee
Robustness. Could you define what you hear by robustness?

Interviewer
Yes. Um, there's a lot of definition. Like there's a different way to define robustness mm-hmm uh, but I can give you one example. Um, for example, you have a, a two, a model that predicts whether the animal in the picture is a Wolf or yeah.

Interviewer
Let's say a dog and your picture of Wolf are always in the snow background. And then you move this Wolf to a, let's say the desert. Then it'll not predict Wolf because Wolf, because it's used to see a Wolf with the snowy background. So this is like, mm-hmm, a example of robustness.

Interviewee
Okay. That, that sounds, um, we haven't faced it, but it actually makes me wonder whether.

Interviewee

If we change, like I'm thinking about in our particular model, what happens when we do the prediction during another month? Like during another season, like, are we still gonna see Christmas recipes in June or. Um, and it's very uncertain for us because of the lack of data. Like we can't experiment with it and like try things out to know if that is an issue, because right now, a lot of what we have is the data is very similar.

Interviewee
The interactions from the users is very similar. That can make sense with what we have. Like it's summer are when we get like ice cream recommendations. I don't know how that's gonna change once the context change.

Interviewer
Okay, thanks. That's interesting. Um, have you ever investigated the explainability of your train models?

Interviewee
Have we investigated it in the sense where, how we played with it? And, uh,

Interviewer
I mean, the, the question is vague by definition. So for example, you mentioned. You were trying to understand your personalized surface. Um, so you investigate the, the explainability. Well, you're trying to understand what happens

Interviewee
and yeah. That in that case then yes, we're really trying to understand the, the results of our model, um, by way of lots of experiment and testing.

Interviewee
Yeah. Okay.

Interviewer
Perfect. Thanks. And is there any other quality issue in ML system that you have experienced and that we did not inquire about in this interview?

Interviewee

I can't think of obvious ones. I feel like the obvious ones were definitely covered. There's nothing that comes in my mind. Right.

Interviewer

Okay. It's perfect. Thank you. Uh, I mean, I see you're looking at the screen. Do you want to, uh, is there something you want to say?

Interviewer

okay. Okay. All sorry. I'm listening. Okay. Perfect. Awesome. All right. Um, so I have two, three last question. Yeah. Um, do you have any AI project that took longer than expected?

Interviewee

Um, actually I, I can talk a little bit about not this project, but like a project in the past where we had a timeframe to be. We were trying to predict, um, traffic on certain trains.

Interviewee

Um, like not trust, like the number of people who would, um, board a certain training, things like that. Um, bef we, we knew we had like a deadline of about, we only had six months to produce it, but because of like changes in scope and misunderstandings with misunderstandings of like what was expected, it took way longer because you have.

Interviewee

I don't know, move, um, move away from a certain model and to another, because at first explainability was not an issue and now we wanna explain it better and stuff like that. So, yeah, I think AI projects can very easily, um, take more time just because unlike just like software where we have, at least that's how I view it, where we have a list of steps and we have an idea of how it's gonna go.

Interviewee

Machine learning is. And AI is generally just like experimentation. There's a lot of experimentation that you can't necessarily budget for because in next steps might usually depend on like the, the outcome of an experiment. And maybe you have to like change plans entirely. It's not just like a piece of software where yeah.

Interviewer

Yeah. Understand. It's a, it's a good. Thank you. And did you have any AI project that have been aborted because your team could not, could get good result with the, the given day assets?

Interviewee

No, I, I have not, but I know it happens a lot. I see people frustrated with it, but yeah. Okay. Well,

Interviewer

that's a, that's a good thing that you, it did not happen to you yet.

Interviewer

So,

Interviewee

eh, I know it will. I think it's part of the game.

Interviewer

Yeah. All right. The last question. Uh, so in, oh, sorry. I will rephrase that one. Um, what is the biggest pinpoint in your daily work day that you would like to go away? Pain point,

Interviewee

sorry, pain point. Yeah.

Interviewer

Um,

Interviewer

Okay. The, the first thing that comes to your mind, really, you say, oh, I would like this one to go away.

Interviewee

Just be having to be like, um, it's not necessarily just like AI, just because it is AI with, uh, the fact that we are actually doing clients and people who need the results, just, um, not having to worry about.

Interviewee

Oh, and to explain this time to this client, because now I spent more time on this. I need to be able to explain that. Cause I feel like it puts a lot of pressure, like, um, like in getting those results out results out, maybe because AI is not quite as understood by all clients yet. Like, uh, it, it might be a very blurry thing for them.

Interviewee

They don't really know what the implications are. So having to worry less about that would be helpful.

Interviewer

Yeah, I see. I see, um, like planning the project is difficult and making understand why you have some hurdles. Mm-hmm is challenging too. So yeah, I see.

Interviewee

Definitely.

Interviewer

Yeah. All right. Well, thanks a lot, uh, for the interview and that that's all for me.

Interviewer

So, uh, yeah, we it's really insightful.

Interviewee

Thank you.

Interviewer

Yeah. Thank you. So.

Interviewee

Well, all right. Thank you. Good night. You guys .

Interviewer

Yes. Anyway, I'll see you at movie maybe.

Interviewee

Yeah, maybe if you come by. All right, bye.

Interviewer

Thank you again for your

Interviewee

time. Have of course.

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