Interview 3 - Rached

Interviewer 2

So you can go ahead.

Interviewer 1

Okay, perfect. So what we're doing in this study is we're trying to develop a catalogue of software, well, quality issues and machine learning software system. So basically any issue you might you might encounter with a machine learning software system - I mean quality issue. And as machine learning software system is a system that has a component that is machine learning. For example, you could have a recommender system like Netflix, it has a recommender system. Well, this is a machine learning software system. Yeah. So any problem you might have - you can think of, for example: if the prediction are not explainable, that might be a problem. If the prediction do not hold in a new environment - it is not robust, that's a problem quality problem. If the model is too large, consume too much resources, and it's difficult to handle, that's a quality issue. So any problem in the quality of a MLSS. And the goal of this is to guide you to work on improving the quality of MLSSs. So if we know there's a problem, we can fix it. So we already asked for your permission to record the interview. And basically, we are going to start with asking you some background information. And then we will follow up with basically the core of the interview. And the core of the interview is going to be basically we are going to attack each phase of a machine learning pipeline. So not pipeline, but stage like the development of the model deployment of the model, maintaining the model, etc. And if you have knowledge of some issue you might have encountered in some phase, that's good. You can mention it. If you don't, it's alright, we are going to pass it. And we do not expect you to have experience in all the phases. So

Interviewee

just a quick question. Are you interested more in, in this science perspective of this are also all the engineering part?

Interviewer 1

I guess both - interviewer 2 what would you say?

Interviewer 2

Mainly engineering. But I'm we are not we are not… We are eager to explore scientific issues with you can include anything, but mainly engineering. But if you think that this is something important should be mentioned, just mentioned, there is no problem.

Interviewer 1
So what is your current position?

Interviewee
Well, my main role is AI Delivery Manager. I'm sorry for all the beeping. Sorry, I'm just going to close this. So yeah, I am AI Delivery Manager at Company X. So I'm taking care of all the aspects of project management, backlog grooming all the scrum mastering. And you call that like account management?

Interviewer 1
Okay, super. Thank you. And how much time? Have you worked with AI? Is it? Yeah, with machine learning?

Interviewee
It's not an easy question to answer. I'm not I'm not myself a data scientist. I have. I have a background in more of optimization. So operations research. I used to develop there. And then I did a few things. I would say, four years ago, I moved to AI. For the first time we moved, I worked for a company that did AI as first mission. And then I have been at Company 1 for nine months now. So I would say that my experience at Company 1 is more hands on like it's really in the daily building development and deployment of AI solutions.

Interviewer 1
Thanks. And I know the answer, but I will still ask the question. For what purpose to use artificial intelligence at your company.

Interviewee
It's not a short answer to give because we have we're doing consulting so every project is different projects can lapse from anything between two months and a year and a half. So they have scale and scope that really vary. We've mainly used for forecasting, like projections, predictions, sorry on demand or things like that. We

had a few projects in NLP as well. So speech to text, also documents annotation. What else? It really varies from, from, from a client to another.

Interviewer 1

Okay, great. And so I guess the type of data you're playing with, as you might have to tabular data for the forecasting. I guess you also have text since you mentioned NLP. Is there any other type of data you …

Interviewee

Yeah, and just I'm gonna, I'm gonna, I may have a hard time answering sometimes, because it's on all different mandates, so different problems and different models and different solutions to those problems. But in general, we have. We've dealt with documents, both structured and not. We've dealt with things that look like the timeseries … we're dealing right now with … we're trying to predict attendance in the attractions. So we're dealing with, obviously, for weather forecasts, both the past predictions and the past actual. Other than that, we're interested in dates as well, like the school calendars have impact on the predictions we're trying to make. Holidays, as well. So I don't know how you classify that.

Interviewer 1

Yes, you okay. You use date for? For to predict what sorry?

Interviewee

We were interested in the holidays calendar to predict the sales in a grocery store, for example, oh, okay to get interested in knowing like, is this Christmas? Is this Valentine's Day? Or is this?

Interviewer 1

Okay? Oh, that's great. And then. And usually, after you collect your data, your client gives you your data, and you do nothing from there, or you have to collect more data.

Interviewee

It's a mix of many things, most of the time, the client will need to produce data. And then the level of quality and structure also varies. Most of the time, we'll have to fetch other data, like the weather, the weather data, we need to obviously get it from an external source. In a case, I have in mind, also the calendars. So school

calendars, we need to figure a way to get them in fact. So it's not something that is publicly available easily. And it depends on Geo localization and things like that. So to answer your question, there's always a part that comes from the client. And there's most likely a part that comes from our own gathering of data from public sources or, and or creative ways to get there.

Q2.1

Interviewer 1

Yes, I'm just curious, how do you get the like the holiday dates, for each country ?

Interviewee

so far, we've dealt only with Country 1 holidays or and let's say Country 2. So that one is easy. Those are those lists are easily available. We have not I personally have not been in a project where and when we had to deal with internationally. The school calendars are fun, because it depends on the school board. So you have to get that from from the various ones.

Interviewer 1

And you do it manually, I guess.

Interviewee

So far. Yes. And then the project I'm working on is really at the really early stage. So we're doing basically just a proof of concept. So manual gathering of data is okay for now.

Q2.1

Interviewer 1

Okay. So moving on from the data, which model is used usually? Is it like deep neural network or it's more linear regression? If you know that?

Interviewee

I don't have those answers. I would say I only understand the overall picture, I do not know the exact models and algorithms and approaches that are used them that you cannot say them.

Interviewer 1

And do you know where your team deploy their model usually? Like either Azure, AWS or Google Cloud?

Interviewee

A bit of all of this is a mix of either the clients infrastructures or our own ones. Okay, so sometimes,

Q5.1

Interviewer 1

yep. Okay, and usually what, when it's your own ones, which one do you prefer?

Interviewee

Good question - we usually adapt to whatever the client is used to, because we work hand in hand with the people at the client site. So they have also to do some maintenance. So we'll we'll adjust with the preferences of the client. I'm not aware of preferences. So I don't know.

Interviewer 1

Okay, thanks. And yeah, so that's it for the first questions. Now we'll move on to the body of the interview. So I'll start with a general question, what are the main quality issues you have encountered with your data or your model? So far, like when you think of quality issues in MLSSs, do you have one case that comes to your mind?

Interviewee

And again, it's a high level view, is that data, I would say is never as expected never like 100% as expected. So depending on the sources of the data, like, for example, we had a client that had free text fields, for addresses and things like that. So the data is there, but it's not structured, where as it should be, or it would benefit from being. So we usually have to go through many loops, sometimes loops or required and to get some data that works. So conformity of the data like is not as expected? Yeah. Sometimes it's the quantity that is not there. I don't know if that's in the the quality issue category. But on many mandates, we were under the assumption that we had sufficient data, and it wasn't the case. So either that was the data was not there, or it was an insufficient quantity. And it's more of a business context or something where maybe it's a misunderstanding of the clients. I don't, I would say the client understands, like, a lot of data is this, but we were meaning this. So we've it's also a misunderstanding on their side.

Data-low-quantity   Data-wrong-format   Q1

Interviewer 1

Did the client gives you like 10, an example or one hundred examples or?

Interviewee

with with the with the quantities? Document a notation we had to do. So processing a document and being able to say - let's say it was a law document in there was somewhere a date that said, this law is enforced as of this date. So we need to figure the date. It needs well, I was going to another point, it needs a annotation. So the absence of annotation is also a big issue. The absence of annotation that are not accurate. So in the case of this customer that had to annotate documents we were expecting, I think it was 100 and something documents - because we were aware it was hard to annotate because a human had to go through and read the thing - and then we ended up having something like a few a few tens. Like I think it was something like either 25 or 30. So that that's really not a lot. At 100 and something it was a challenge - we knew it would be a challenge -but then it was even less than that.

Data-low-labeled-data   Q1

Interviewer 1

Okay, and when do you meet annotate is it like to give a label to a document or is it to say this word means this, this word has this label, etc.?

Interviewee

it's really giving labels or … sometimes it was labels, sometimes it was meta information. So associating this document, the law explaining this document starts at that date, it ends at that date, it's that kind of category of law, it applies to those countries to those type of something, something. So sometimes it's a label, it's like a category or something. Sometimes it was a meta information, like a date. Okay.

Q1

Interviewer 1

Okay, see? Okay, so, I mean, we're gonna move on to the next section. I mean, the next section is pretty related to what you said, but I'm gonna go over it, and maybe we're going to find new things. So it's about data collection. So I will go through every data collection method. If you think of another one, just tell me and

basically, I'm gonna search if you ever had problem with each data collection method. So; did you ever had like a data collector, so you pay someone to collect your data for you?

Interviewee

I'm gonna answer for Company 1's perspective. No, I don't know if clients did that. But I'm not aware. Okay.

Q2.1

Interviewer 1

Well, this one is, yeah, this one that I know you did. Do you sometimes use public datasets?

Interviewee

ohhh yes, the answer is yes. Sometimes you're going to use it when either the client's data is not ready. or it's not available, for example, we were doing dynamic pricing. We know the past data does not fluctuate. So there is data, but it's not suitable for the need we have. So we're searching also for public data. We are often looking for publicly available challenges' datasets.

Q2.1

Interviewer 1

on Kaggle, for example, that asked for a challenge. Okay. And do you have any quality problem with these data sets?

Interviewee

I don't know. Because I haven't dealt with them enough. Yeah. It would be it would be okay. Right.

Q2.2

Interviewer 1

Maybe Maybe. I don't want the I already know us weather service API, but to use any, any other type of API in your company.

Interviewee

Whether is clearly the most, the one that is used them in most projects. If you're thinking about API's, and I cannot see anything else.

Q2.2

Interviewer 1

Is there a problem with the this weather API, you have at Company 1?

Interviewee

There are many providers, right? So I'm trying to think - I am not aware of major issues. Sometimes the data we're looking for is not there, for example, like the past predictions, not the past actual but the past prediction, like at that date, we were forecasting, sunny with a few clouds in seven days from now. So having the past prediction, sometimes is not available.

Data-missing-features   Q2.1

Interviewer 1

And do you sometimes have a cascading model? So one model predicts and creates training data for another model? Do you know what I'm saying? Like do you have one model that create a prediction, and it becomes a feature for another model.

Interviewee

I see what you mean. So do we use such processes?

Interviewer 2

I think the question of the first interviewer is that did you use machine learning for other models? or other purposes?

Interviewee

That is the question. I understand the question. I'm trying to think, but I'm not seeing anything where the result is used as a feature. It can be used in the process down down the line, but as a feature for another model? I don't think so. Okay. It's nice in the projects I've been I've been involved with.

Q2.1

Interviewer 1

Yeah. And so have you encountered any other data quality issues caused by data collection?

Interviewee

think we covered like the absence data, the absence of labels … I am trying to think of anything else than that. Oh that's a funny one, we had data that changed. So at a pivot moment, the data was not only structured differently, because that's not necessarily an issue, but it was, by lack of a better word, it was organized in different ways. So we had challenges adapting the data from now to the data the past data that we needed to use to make predictions from there, I don't know if that makes sense.

Q2.2

Interviewer 2

Can it be formulated as a consistency issue?

Interviewee

You see, in that case, and the case I have in mind, it was a business decision. The client decided to organize the products they were selling in a different way. So you're like okay, now I'm selling I was selling ABC now I'm selling DEF but there's, there's a mapping that can be done, but it's not obvious sometimes that we're grouping some products together. So there's also a challenge of reconciling meaning with the past data. So I would say more of a an issue of data in time

MM-drift   Q2.2

Interviewer 2

according to your explanation, they have not converted the old data They were keeping it the original way, and now they use some translation. Okay. I see.

Interviewer 1

Yes, I think I know what you mean. Did you talk to me about it? Is that your like your client changes … like he changed his deal to his clients and this change like the function to predict or something like this, because I'm trying not to say the name of the company.

MM-drift   Q2.2

Interviewee

yes me too, I don't know if we're talking about the same thing.

Interviewer 1

Okay. All right. Well, thanks. Thanks a lot. So we'll move on to the next question. We're in data collection and we'll move on to data preparation, maybe you will have less knowledge of this section. There's no worries. So data preparation, it's split in two parts, you have data cleaning - so fixing errors in your data set - and then you have, like, putting your data in a format that is good for a machine learning model - so data transformation. So I will start with the first one. Do you know which tool your your team use for data cleaning or data transformation when they prepare their data?

Interviewee

I couldn't say.

Q3.1

Interviewer 1

Do you know what are the pain points that your team repetitively encounter when they prepare data?

Interviewee

I'm trying to think. Most of the time, we don't we don't I don't think we see things coming. Like we know that there are going to be issues. But it's always like, it's always a bit of a surprise, like, you open the hood and then you see the data and you go like, Okay, we were not expecting this. I'm trying to think of recurring issues. I'm not too sure. But I would say sometimes you don't have the data. Like there are like there is a prediction that needs to be made on something precise. And we need a feature for this. But we don't have the data. Neither directly or we cannot deduce the information from the data.

Data-low-quality   Q3.3

Interviewer 1

Thank you. So yeah, I think that's it. Is there any other data quality issue we missed in general, and you consider relevant? It's a large question.

Interviewee

yeah, no, I think we covered the main ones. I don't I don't see anything else for now.

Q3.4

Interviewer 1

Okay, perfect. Thanks. So we'll move on to the model part. So the first question is, how do you how does your team evaluate the quality of your, of the model you're producing, how can they assert that this model is good enough for the client?

Interviewee

Right? Most of the time, obviously, we have the ground truth. So we can test the predictions or whatever, like the results against ground truth for this sample we have. On every project or so we'll also have a baseline. So we'll implement something else, like a naive approach or something, we're going to automate whatever the human is doing right now. Or, we're going to find out like a heuristic to try to answer the same question, and then we're gonna, we're gonna compare with this. So in other words, is the full blown ML model is worth the efforts compared to a naive approach or something like that. So there is the pure accuracy measure that we can do if we have the ground truth. But there's also is it worth the effort? From a business perspective, from a ROI perspective?

| ME-metrics-a-are-poor-proxy-of-quality | | Q4.1 |

Interviewer 1

Yes, I see. I see. It's a baseline basically. And so do you have any user acceptance measurements? So you, you have your model, you go to your client, and they try your model? And they see if it's enough or not enough for them?

Interviewee

It's funny you ask, because from, from very technical and sciency people, sometimes we managed to have very clear criterias and everything, but I realize more and more that from business, people that know their business, like they know if the answer is good or not, from the feeling, it's really, really hard for them to to try to come up with a number. Like, if it's accurate, at 95% I'll be happy. Whereas maybe if they say it's a 83, but then the rest is to very complicated or very, like non trivial cases, I'm okay with that. But giving numbers like trying to see, I'm going to accept, quote, on quote, the solution if it reaches this level of accuracy. It's really hard to get from business people.

| ME-Metrics-Meaningless-For-Stakeholder | | Q4.3 |

Interviewer 1

I see. So it's for the criteria for success, like in numbers, but so you have to meet the client and then check with him if it's

Interviewee

yep, that's, it's really fun because we it's funny because we did that like last week for a major project we delivered. And it was basically that, like we produce predictions on a given period that we agreed on that was sufficiently complicated and everything. And we sat with the client looking at things and producing statistics, and things like that. And what most of time happens is okay, but what happens with in that case, it was products. So predicting demand on sales, sorry, sells on a given product. And they'll go like, Yeah, but this product in this in this store, what happens with them, like, Oh, this is not like this is way off or something. So they know their business, that they're able to challenge the results. But the right way to address it, I have a feeling it's more it's not through statistics and mathematics and bounds and things like that. It's more like, do you have a good trust in the results from your intuition? Which is in a, when you have a very elagant equation in mind, it's hard to say, this is gonna, this is how they're going to accept or not the solution, but sometimes it works better.

changing-requirements   Q4.3

Interviewer 1

Yes, it must be a really stressful, stressful time. You build your model? And at the end, you check with a client?

Interviewee

Well, you don't wait at the end first. Yeah. Yes.

Interviewer 1

So you didn't meet them more often? And more often, you can,

Interviewee

I would say, it's really important to have something like someone who can challenge their results and who get it, who knows the business, like, deep, and that can that can guide your your work there. So it's your it's important to involve them as early as possible.

| how-to-evaluate-model | | Q4.3 |

Interviewer 1

Thanks a lot. Sorry, for the pause. I lost track of where I was. Okay, I'm back. Are your models sometimes used in scenarios that involve different groups of people?

Interviewee

Okay, I got I got an interesting one for you, I think. And let me know if I understood your question correctly. But we're doing predictions of sales for products in the grocery. The fact that a product is on sales like promotion on that week, it really has an impact on the sales. So we discovered through the project that we needed to treat these differently. So in the group of product that we needed to produce predict predictions for those that are in promotion behave differently than those that were not. So that's one example of like one silver bullet was not enough anymore.

| Data-metadata-issue-fix | | Q4.4 |

Interviewer 1

data if you had to know it to understand why prediction, sales were better for a product at some point in time,

Interviewer 2

I think you're talking about the most influential feature.

Interviewee

Absolutely. I see what you mean, there's there is that and there's also the fact that we discovered in the same in the same analysis that the products that had very, very little sales in terms of units, they were behaving really excuse my French, but crappy ly with the like the big ML model, because it didn't make sense. Like you sell that thing, like once a week or once a month, you don't have sufficient data, but yet we were processing them in the same way as the other products. So that really made us go toward that I wouldn't say specialized,

but we had to adjust the approach to those two, those two different things. When I think of it, the being in promotion or not, is more of a feature, but having two little historic data, then that's another problem. And that's that's more of a difference in the pool of, of products.

Data-low-quantity   Data-low-quantity-fix   Q4.4

Interviewer 1
Interesting, and are there any other quality issues during the evaluation of models?

Unknown Speaker

Are you interested in scaling?

Interviewer 1
Absolutely, yeah,

Interviewee
absolutely. Yeah, that was that was one of the main why it wasn't an issue because we knew it was. Let me let me rephrase this. We made decisions early in the process in a given project, knowing that we would have some work to do to scale it. So we've reached the point where the machines and the approaches and the libraries we were using, they were no longer able to cope with the amount of data that we wanted to process. For example, while predicting sales in grocery stores, we were dealing with a bunch like a handful of stores. And then we were we wanted to scale to like the whole division. So like, multiply that by 20, the current approach was no longer viable. So we had to do a lot of work to be able to scale that in a production environment that would hold it.

Q4.5   QM-scalability-issues

Interviewer 1
Okay, so when the problem was getting more complex, simple model, simple models are not enough you have to have,

Interviewee

I would say, That's why I asked the question at the beginning, if it was more engineering, or science, the model was perfectly fine. So the main model we use for this was fine. But all of the deployment technology around it needed, needed to be changed. So it was really a data engineering and software engineering challenge.

MD-issues-with-dependencies   Q4.5

Interviewer 1

Okay, the infrastructure had to be changed because of data came from different places and something like this?

Interviewee

And the data was pretty much the same, but the quantity of data. It was multiplied by 20 ish.

Q4.5

Interviewer 1

Okay, and the pipelines were too slow?

Interviewee

Pipelines and all the all the processing steps.

Q4.5

Interviewer 1

All right. And let's move on to the Model Deployment question. So we already know where any models are deployed. Usually, when you do deployment model, do you have a pipeline to deploy the model automatically? Or you manually do it each time?

Interviewee

Depends on the client depends on the the client, if we're deploying on their infrastructures or not. Some client will be really manual. Well, first of all, the best practice is to have CICD and to have like, automatic deployment. That's what we push for it. Sometimes it's impossible. So sometimes we'll have clients that will keep doing it manually.

Q5.2

Interviewer 1

And it's impossible because they're not interested, or it's impossible, because they're too, like, complicated for them.

Interviewee

a bit of both sometimes they don't have the bandwidth to put in place what's needed. So it's a mix of our I'm sorry, bad reasons.

Q5.2

Interviewer 1

I see. Thanks. And did you ever have a model that locally when your team, build it locally to perform great on data set, but once you deploy it, it was not as good anymore?

Interviewee

Not good anymore. No? Like, like accuracy and performance was comparable, but we had, we had at least an issue where the results were significantly different. And we had to investigate, because technically, the inference part should not differ from or was it the inference part was it the training as well? Anyways, we had, we had a few issues like this, I have a feeling I don't know the details, but I have a feeling that we build things so that this, this did not does not happen. But in that particular case, it was something as stupid as in a sheet in the random seeds. It was it was as stupid as that. Like it's in a local environment, it was initiated, when not initiated, it was initiated in a way that was different than production. So the fact that was not explicitly initiated. It was taking the default behavior, which was different from production than local environments. Okay, so that was a, it was a silly mistake. But other than that, I don't think we have major differences from local to production. It's pretty from my understanding, it's pretty rare that data scientists or data engineer will work locally, like purely locally because the machines just don't cannot handle it.

MD-configuration-issue   Q5.3

Interviewer 1

Okay, perfect. And you were talking about random seeds. So I guess you're you retrain your model on the cloud, before before deployment,

Interviewee

or even like during production, because, for example, the predictions in the grocery stores every day, it needs to be retrained. It is not trained once, because every day we have new data, we have new historic data. So we retrained and then we do the inference.

Q5.3   reactive-fix

Interviewer 1

Okay. Okay. So the problem was not that the data wasn't good enough, not good anymore. You use new data each time, but it was that your model for some reason, didn't perform as good on this data set for training.

Interviewee

It was same dataset, same model, same everything. Well, sorry. It was being retrained. But it was exactly the same data set and behave differently. So that's what was suspicious does

Q5.3

Interviewer 2

the random seeds normally doesn't affect the data, but the model for the training process?

Interviewee

the training part, because I think there was a some exploration of parameters that was based on I can't remember, we usually fixed a random seed. Yeah. So that we avoid having, that's it. So we avoid having non deterministic results. But if you fix it to something that is different in the two environments, then you have different results.

Q5.3

Interviewer 1

I just have a question. So you retrain the model every day, but you used new data to retrain a model, right? Okay, perfect.

Interviewee

Well, it's part new data and part old data. So let's a sliding window, you train with the past two years of data. But every day that that window goes, further to the right.

Q5.3

Interviewer 1

And is it sometimes difficult to detect this? Like, it takes some time before figuring out that a model is not good anymore?

Interviewee

That's a good question. I am pretty sure there would be tools and monitoring, monitoring tools in general that we could put in place, but most of the time, it's going to be your customer that will say like, Hey, we received this today. It doesn't make sense. Yeah, I see. A human that says … but we do some monitoring. Um, there's just not aware of the details.

how-to-evaluate-model   Q5.3

Interviewer 2

In that project, since you have some new data, you prefer to retrain the model?

Interviewee

In that case, yes, yes. Sometimes we will not sometimes we'll wait to gather up new data. And then we'll retrain. In the case of this customer, it's everyday.

Q5.3

Interviewer 2

All right. All right.

Interviewer 1

That's it for model deployment. And maybe the last question as you think other quality issues.

Interviewer 2

So the deployment part.

I would say no. It's pretty predictable, let's say.

Q5.4

Interviewer 1

I just forgot a question. I'm sorry. Regarding your problem for model deployment i asked you if you ever had a model that perform well, locally, but once deployed, no. When did you realize that you have it was

Interviewee

It was simple as explicitly initiating see the random seed to something that was controlled. Because the problem was that we were, there was a, since there were no explicit initialization as we were falling on the default behavior, which was different. So setting in explicit initial initialization is that it resets all the wrong.

Q5.3

Interviewer 1

So you fixed a bug basically. That's it. Thanks. Alright. So moving on to maintenance question. Do you I mean, we're almost finished. I see there's three minutes left, so it doesn't go fast.

Interviewee

Don't rush it. We reserved the time for an hour. So it's okay.

Interviewer 1

Okay, thanks. Okay, I know you deploy your model, generally, and when you do, once a model is deployed, you have to monitor its performance. We talked a little bit about it. Are you sometime responsible for the monitoring of your ML models?

Interviewee

Depends on the client. Again, if we have clients that have people with the right skills to do it, they do it on their side. Otherwise it falls on us.

Q6.1

Interviewer 1

And one it is on your side, how do you monitor your models?

Interviewee

I'm a bit ashamed to see that I don't know. Oh, I think there's two wing. But I'm not sure we put in place or not. Sometimes there are also some alerts. I mean, first I think the the engineering part, the deployment part is easier to monitor. For example, if you have an error somewhere in the in the pipeline, and it's easy to send an email, we'll send an alert or something. The quality of the models, if it deteriorates in time, that's a bit harder to monitor there. And that's the part that I don't know how,

Q6.1

Interviewer 1

if you know it, which metrics does your team monitors?

Interviewee

Probably the same metrics that we look at while building the solution with the client. So sometimes, sometimes it's accuracyif it's forecasting. Some other times with NLP project where we did speech to text, it was the word error rate. So sometimes, it really depends on the the the problem itself.

Q6.1

Interviewer 1

Yes, yes. I mean, that's maybe an obvious question. But what happens when a model is staled? Do you have for example, do you have a platform to automatically return and redeploy model?

Interviewee

I'd say no, we get alerts for sure. The I have a feeling is pretty manual.

Q6.1

Interviewer 1

Do you have models that goes stale after some time?

Interviewee

I'm trying to think and the projects I was involved with, not really, but I've heard of other projects where the predictions were going everywhere. But I have no details. And so

Q6.2

Interviewer 1

Thanks. And do you have any any other issues you can think of?

Interviewee

they're their burly two challenges and that's more gut feeling and experience. But I would say the scaling part of it is hard to to keep in check. Especially if, for example, this this prediction on a window of data for two years, it used to be unbound. So if you're continuously adding data to your to your training dataset, then then it will easily explode from from resources like machine resources perspective. So the scaling part, like building something that will hold it until, like in all cases, as the data increases, the quality of data increases and things like that. There's that but there's also you're never protected from business changes. So if you train your things, and you tune your things with a good understanding of the business and things like that COVID is probably a very good example of this. Like it hit everywhere, it changed behaviors everywhere in all domains and everything. We could not predict this. So that's an extreme case. But at the business level, scale, you can also have that you can have drastic changes, introduction of a new product, introduction of a new way of doing things that can screw you predictions. So there's that it's more like, it's not from a technology perspective, it's more from from the customer's way of doing things, or the user way of doing things.

MM-drift   Q6.3   QM-scalability-issues

Interviewer 1

Thanks. I guess that's the last question last section.

Interviewer 2

Okay, the second one.

Interviewer 1

Is model robustness something you often encounter when building a model?

Interviewee

I would say no. Because most of the time, we're building a solution for a given use case that we don't build on top of this. After that I could imagine easily predicting cells and groceries, depending on the department. As you add departments, they will behave differently. Like the meat department behaves differently than any fresh, fresh department behaves differently than the other ones. So I could see a challenge in introducing a

new department because you have goods that don't sell the same way you don't sell canned food the same way you sell T bones. And canned food versus weather, virtually no impact, whereas the bones if it's raining, or if it's Christmas, he don't sell them. So I could I could easily project that those will be a challenge. Like for him who absolutely need to retrain.

changing-requirements   Q7.2

Interviewer 1

And finally, as you ever investigated the explainability of your trained model?

Interviewee

not in depth? I'd say. I'm trying to think it's, and by the way, it's really how would you say that like it's really important for end users, especially if they're not from the data science domain. For business owners and business people to understand why this machine said that. It's really important. It comes in every discussion we have with with clients. Do we try to do something? I'm trying to think.

Q7.3

I would say no, not really. But my my intuition is that we have a good idea of what to try and what to do, but it has never faced actual production like being able to explain things in a business context.

Q7.3

Interviewer 1

Okay. But it's sometime it's important for your client, not for production, but they would like to know

Interviewee

most of the time and it needs to be like lamest terms, if it's the gibberish, the mathematical gibberish behind like, it doesn't mean anything to them. So the challenge is also to make it intuitive and useful for them.

Q7.3

Interviewer 1

And it comes back to what we asked you earlier on about user acceptance. When you go see the client, and you ask if it performs well.

Interviewee

In extreme cases, I would say, like something that is completely off, they'll go like, Okay, why? The cases that are that are spot on, they don't they don't like, get on discuss them.

Q7.3

Interviewer 1

I see. Thanks. Yeah, is there any other quality issue you can think of?

Interviewee

Oh, hold on. I had an interesting case, in this. Okay, here's one, document and annotation. So again, we need to figure the dates of application of a law that is described in an unstructured way in the document. We realized too late in the project that, in fact, even a human could not do it. So part of a sample, part of the data set that we were using was, in fact, not even processable by a human, I would give this document to this person. And the only way they had to figure out the date was something as crazy as the date of creation of this file or the absence of another file that supersedes this means that something else, so information that was completely out of the document itself, because the assumption we were we were making is that the information is in the document. So we had cases where predicting something or classifying a document was not even possible for humans. You needed business contract context, or meta like meta meta information around the data.

Data-no-signal   Q7.4

Interviewer 1

Really interesting, so it's a mix of maybe your clients have very high expectation of what you're able to do. And also there's missing data.

Interviewee

Something like that. Yeah, it was done. Like, it's even impossible to get the ground truth without knowing external information, like the date of the creation of the file itself, like the PDF file, or relating documents between them, like knowing that this document exists, that means that this one is no longer an action, and I cannot make the link there.

Q7.4

Interviewer 1

Okay, so this is a bit out of the scope of the interview. But I have a question: do you often have to tell your, like your client, that very high expectation, you have to lower them? That's what happens when you started, because most

Interviewee

of the time, people will be understanding. Like, once we discovered issues with the client, they understand. I mean, it's hard to argue the fact that like, you're asking me for the color of something that has no color.

Interviewer 1

Actually, I have two more question. Three more questions. And then

Interviewee

how many last questions do you have?

Interviewer 1

Every time I think I'm finishing i have another section, but it's real. There's really three question. So do you have any AI projects that took longer than expected?

Interviewee

Yes. Most of the time. It's pretty … I cannot remember of a project definition ahead of time. At best you're out in the estimates we're in terms of timelines and everything unless you have a restricted line. In those cases, we can do it, but if you don't have a strict deadline, it is often late.

Q8.1

Interviewee

You sound like a few of my colleagues that challenged this. Well, obviously under estimating the the effort sometimes the data gathering and preparation phase is much longer because we wait for the customer, sometimes we also have surprises, approaches that we think that we're going to work well don't. So you're like, Okay, back to the drawing board. So it's a mix of all of this anything I say?

Q8.1

Interviewer 1

Okay. And did you have any prior project that have been aborted?

Interviewee

Almost, I would say. It's hard. It's hard to get a project that you've been putting your blood and sweat. And so it doesn't happen often. Should it happen more often? I'd say, yes. We did it, I think, once.

Q8.2

Interviewee

Not too sure. But I have a feeling that both the client and the people that worked hard on a project, they all have bias because you put so much effort in there that you want to bring that to production, because because you worked so hard on it, and you spent so much money there. So it's rare that we kill the projects, it's sometimes it's really good, because continuing and blocks things later on.

Q8.2

Interviewer 1

So when your team is building a MLSSs, which pain points would you like to go away?

Interviewee

That's a really good question. All that comes to mind is like the uncertainty of will it work or not, maybe speeding up the process? How can I say that getting to a point where you can assess the quality of an approach faster, that would help. Sometimes also, it's downtime due to training time. I remember projects where the training was like three to three to four hours now. It's a bit long to validate something. If you do something in the morning, then it was part of your day. That's what I know. There are also concerns about having clean data faster. I realized my focus is really on getting things faster, like less effort in time to get to a working point.

Data-low-quality   Data-low-quantity   ME-testing-model-challenging   Q9   QM-efficiency

Interviewer 1

That's true since your manager.

Interviewee

yes, delivering the right thing fast.

Interviewer 1

So thanks a lot for your time. It was really, really interesting. And I think you underestimate how much you could contribute to the study.

Interviewee

I didn't know what kind of question you would ask. But I agree with you. It was - I could answer more than I thought.

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