Interview 10 - Rached

interviewee

Thanks. This meeting is being recorded.

interviewr

All right. Uh, yeah, go ahead. Uh, so can we have some background information about you, Interviewee? How much experience do we have? What is your current position? Anything?

interviewee

Uh, yeah, so I have, so I have other experiences, but I'm gonna focus on, on data science. So I have a master's.

interviewee

I guess I have to give the whole information. I have a background in business. I studied finance for my bachelor. Then I work in finance, uh, for I think two years, three years. Um, and then I switched, I, I went to School 1, said to do a masters in data science. I think it's called business analytics and data science or something like that, but it's basically data science.

interviewee

Um, I did the masters and then since then, like I've been doing, uh, you know, uh, work in different companies by myself also, uh, as. Independent consultant. And right now I am a data scientist at Company 1. Does that answer your question? You want more details? Yes.

interviewr

No, it's perfect. Thank you. Good. Yeah. Uh, what are the main quality issues you have encountered with your data model or systel so far?

interviewee

Um, so I'm gonna speak about a specific project that we did, um, in the last few months. Um, so data, Um, I encountered, we encountered, uh, an issue with an external data provider, uh, in the way that they sent us data. And to be specific, it's, it was weather data. Um, they would like, we would poke their, their API for

different, um, different regions in the world.

Data-unstable-provider

interviewee

And you know, like it returns with weather parameters for a given set of dates and. We, we would do this like on a, on a weekly basis, and then just to be safe, we would poke the API for repeat dates. Let's say like, this week I'm gonna poke the API for the last eight weeks and the next week, the last eight weeks, and so on and so forth.

Data-unstable-provider

interviewee

So, um, all of a sudden we, we noticed, um, so like I'm telling you, well, I'll tell you a little bit. So like it was, the issue was with weather data in the sense that we would get the data. But the data sometimes did not make any sense. And how did we notice? We noticed because the performance of the model started deteriorating.

Data-unstable-provider

interviewee

So we only, So it was, it was with one month delay that we noticed that there was an issue, right? So like we saw the impact in our model came only one month later. So we, we saw the, the, the impact in the, in the model performance. And we were like, Okay, what's the problem? And then when we did the deep dive into the data, then we saw that the weather data that we were receiving, was all over the place.

Data-unstable-provider

interviewee

Uh, and it didn't, didn't really make sense. And then when we contacted the, uh, the, uh, the data provider, like they said that, you know, uh, it was, it was, it was a known issue in their platform that they never communicated with us. Uh, but it was an, it was a known issue, uh, for some of the, some of the problems and then the other problems.

Data-unstable-provider

interviewee

They, they just couldn't have a, an appropriate answer. So, so far so good. Is this, is this answer okay for you?

I can, yes. It's really good. Yes. I can dig, dive into like what we did afterwards if you want, or, uh,

Data-unstable-provider

interviewr

yeah, that was my next question. Yes. Or how do you handle this kind of issue? How do you prevent them

from happening

interviewee

also?

Data-unstable-provider-fix

interviewee

Right. So what we did, again, like, so we noticed this was a problem and then we said, Okay, like we don't

want this to happen again. So our. Reflex was, let's look for a different data provider, different weather data

provider. And you know, like there's, there's many out there. And then we tried to, you know, like better

interview I guess the, the providers to see like, okay, what guarantees can you gimme that data is gonna be

good?

Data-unstable-provider-fix

interviewee

Have you had any issues? Like, are you gonna communicate any issues if they ever happen? Um, so. We did

this process, we selected a better candidate, but still we had the question like, okay, what if something

happens in the data? And, but it, it fucks up. Like, sorry, like my language, but like, you know, it f up my, my,

my predictions again, like my, my client's not gonna be happy.

Data-unstable-provider-fix

interviewee

Like, you know, like the, the usefulness of them, it's, it's, it's lost. So what we did is we did, we, we, we loaded

and alternate data source for weather. . So instead of having one, we got two new data sources. And what we

were doing is we were comparing them on a weekly basis. Uh, you know, like with different thresholds, like to

see like if there's like, you know, we would compute the, the, the, the, uh, the differences across time.

Data-unstable-provider-fix

interviewee

I don't remember the exact parameters, but like, you know, like maybe across a month or across a couple of days. And then we would like see the standard deviations, like whenever, like we were above a certain parameter. We would flag them and then we would send alerts to the, uh, user saying like, Hey, like there's a difference between your two data sources for weather.

Data-unstable-provider-fix

interviewee

So like, maybe one of them is wrong. Maybe the the one that you're using the model is wrong. Maybe the other one's wrong. We don't know what's got, what's going on, but just be, be aware that something might happen. Right? So this is what we did for weather, right? So this is what we did for, for this specific source, which is weather.

Data-unstable-provider-fix

interviewee

Does that make sense? Like so we basically compare another one and then. We, we just made sure that like they were following as closely as possible and then whenever they were not like we would flag say like, Hey, like there's something weird going on.

Data-unstable-provider-fix

interviewr

Yeah. It's super clear and interesting. Thank you. Yeah.

interviewr

Um, alright, so I will, I will talk about basically, Oh yeah, go for it.

interviewee

I have more, I have more examples with the other data sources. I don't know if you want me to tell you

interviewr

or Yeah, go for it. Anything else you want to say? Yeah,

interviewee

so other stuff, like for example, we were using public data. Um, that, you know, are it, it, it's updated on a weekly basis.

interviewee

Like actually like every two or three days it's updated and sometimes you get the data and send, like, sometimes issues can happen with like, with our api, right? So for this you don't have any, any, any given guarantees, um, as to like, you know, like that always gonna be, their APA is always gonna be working, but at some point we noticed that it was down and again, like, um, we.

interviewee

We were trying to cover for this by failing whenever we didn't get the data right. So like, it's like we have for, by the way, I'm speaking about a forecasting project. So it like the, the goal was to, to make it forecast on a, and that would update on a weekly basis. So we would have different data sources.

interviewee

One of them was water weather, which I already talked about. The other one was some government data. And so what we would do is like, if it fails, like if I fail to fetch the data, then just fail, don't generate any predictions because you know, my models are trained with data. So if I don't have one source of data, then you know, like my predictions are not gonna be accurate, and so on and so forth.

interviewee

Um, so we were, we were protecting for that except that, you know, like whenever we were not getting the data, then I was not generating predictions and then the client was not able to use that information, like, you know, the predictions. So what we did is like we, we. To bypass that. We, we made an assessment of what fundamental, uh, sources of data were.

interviewee

Um, so which ones we were fine with not getting information on weekly basis, right? So like there's models that are, like the software is built so it can handle nu values. For example, like I poke this data source, I don't have it this week. I can still make predictions and then can handle that very well.

interviewee

So we would, we made that assessment for some sources that we didn't. Consider main drivers for the predictions. And, you know, the other ones would just build the logic for imputation. So like, whenever we would, we wouldn't get it like imput and then keep track of it, right? So like whenever you do something that you're, that is not, um, in line with how you joined your model, then just keep track of it instead of just not drawing predictions, generate predictions, but keep track that you did something different.

interviewee

So maybe. Something's gonna, like ex don't expect your results to be super, uh, precise, let's say. Does that, does that make sense to you? Yes,

interviewr

totally. That's clear. Okay. Thank you. Yeah. All right. So yeah, go, go, continue.

interviewee

No, that's, that's, that's it for, uh, that's for data sources. Um, and then, Yeah, I don't know.

interviewee

Like, like, tell me like, I mean, you're the one interviewing, so like, I don't, I don't wanna, I don't wanna do your

interviewr

No worries, no worries. If you want, you can continue with the same question I ask you for data model or system, or we can continue. We have around 20 question,

interviewee

25 question. Okay. So like for, for model, you know, it's related to that.

interviewee

Um, the issue here is that we have, you know, we have a forecasting problem, so it was known that. The performance of our model was not consistent throughout the year, right? Like, you know, like when you're forecasting you, like you can have whatever you can be forecasting on a daily basis, weekly basis on yearly basis.

QM-explainability-issues

interviewee

Like the, the, the accuracy was not the same throughout the year. Um, so, so for this, we, we built one. Um, that was really like, I guess the main thing was af was what would only allow us to check performance on an after the fact basis, right? We were not able to build something to alert us. Um, even though it would, like, you know, I, ideally you would wanna know like, Oh, something's weird with my, this week's predictions.

QM-explainability-issues

interviewee

Like, either, like, my confidence intervals are too wide. Like there's some, some something that my model like is telling me that it's gonna, it's gonna be weird. But we didn't have time to put that in place just because we didn't, you know, we didn't have basically budget to do it. So we built something that it was more on an after the fact basis that is just literally, like, go back in time and then see like, okay, how, how, how am I doing for, for my past predictions that I've done in the last few weeks?

QM-explainability-issues

interviewee

Like I said, is it consistent with my, uh, you know, the, the accuracy that I was expecting when I trained my model and when I decided to deploy, or it's d. So this requires manual, like manual work. Like someone, a person needs to go into the, the report, if you will, and then see, okay, is making sense. It's not making sense.

QM-explainability-issues

interviewee

But again, the problem is that it's only on after the fact basis. If you already made a decision on that, you're screwed a little bit. Okay, I see. We're, we're, we're bypassing this a little bit because our system is, is for helping. People make decisions, right? It's not to replace their decision a hundred percent.

QM-explainability-issues

interviewee

So like the, the user takes our in, takes our output, uh, as their input, like just to validate like their, what they, like, for example, what they wanna do. Like, okay, like I see the models telling me to do something. This even before looking at the model, I was, I have, I had the same idea. So yes, it's validated, I do it.

QM-explainability-issues

interviewee

Um, so like that's why we are a little bit protected because it's not automated a hundred percent. Like we're not making decision. Right away from the model.

QM-explainability-issues

interviewr

Okay, I see. Yeah, it's clear's pretty clear. I I the last follow up question because it's clear, so yeah. Okay. All right. And the last one, so I ask you about data model and uh, the last one

interviewr

was system.

interviewee

System. Do you mean like architecture wise, like, and like the different engineering components?

interviewr

Yeah, exactly. So, uh, so we are interested into ML system. Uh, you have data that may have quality issues. You have the model, and the, the machine, machine learning software system, like in general, uh, may have quality issues that are related to the fact that you have ML component in it.

interviewee

So, I don't know, like for me, like, you know, like the, the ml like, you know, the ML model either because your data is changing and then you're making predictions that are not consistent with what you have in. Training and validation, let's say. Um, and then of course the, the data changes, like, then everything changes.

MM-drift

interviewee

But in, in terms of system, for me, I'm thinking more of, you know, the engineering that comes around it. Like we have a system that, you know, like triggers every week. It does, uh, the etl, it does model retraining and then it does the inference on a, on, on, on a batch basis, right? So, So for me, the engineering, we did have some, some hiccups just with this, with the system where we're using, uh, uh, Azure.

interviewee

And then, uh, you know, there was some functionalities, I guess like of the system that we were using in Azure that it was not clearly defined. Like if you, if you were reading documentation, like you wouldn't find an answer for the problems that we had. Uh, but most of the time those problems were solved by just restarting the.

interviewee

So like the pipeline was, had a trigger to run every week, and then for some reason it would stay running for hours and didn't make sense. Like when it actually takes like, I don't know, 10 minutes and it would stay, it would stay running for, for hours and it, so it's like it bugged forever and then it was just like relaunching would solve, would solve the problem.

interviewee

So that was, that was a bit of a, Okay. Yeah.

interviewr

Okay. Why did you mention as Azure?

interviewee

Uh, just because we were using, So the, the system was running on. Oh, okay.

interviewr

Okay. I thought it was, uh,

interviewee

Microsoft story. Yeah.

interviewr

Okay. See, And so, so you said that the, the, the ML pipeline was, uh, bugging at some point.

interviewr

It, it run endlessly, right?

interviewee

Yes. Yeah. It could be, it could be a different steps. It sometimes it would happen at the fetching data or pre-processing data. Or sometimes it would, it would bug at the modeling part, like, you know, like at the train model and then make predictions. It, it was a bit random like, but it sometimes it would bug and then it would stay there for hours and then you had

MT-predictions-low-quality

interviewr

refresher.

interviewr

Okay. And this issue, they were not consistent. So did they change from time to time? Yes. Yeah. Okay. And how do you, all these issue.

interviewee

For now. Like, like if they happen Yeah, they, they, they happen. Yeah. They, they, they still happen and like it's someone has to go in and, and relaunch. Okay. Okay, perfect. The, the, the like, I mean, I'm not saying this is the best way to do it, right?

interviewee

Like it's, it's also because, you know, there's, there's money constraints I guess, right? Like the customers are not willing to pay and differently for you to fix box if they deem this is like not very important. Like if someone. Clicking relaunch once a week for like 10 minutes. Then, then they don't mind like, just keep doing it less, as long as I don't have to invest more money.

interviewee

So, Okay. I'm sure like, like, you know, the ideal thing would be like to investigate more and then like, see why, why that's happening. Change the system, like change the approach and then like do it so it doesn't happen again. But, you know, money is a constraint.

interviewr

Yeah, sure. Uh, it, it's a good, it's a good answer.

interviewr

I, uh, I. Yeah, we are not evaluating if you, you have the best practices start out. I mean, what you said is a really good answer for, for the, for this interview. All right. Uh, so, so we were really general and I will ask now, I'll ask you questions that are more specific to each, uh, phase in ML workflow. So from data to the model and to monitoring the model.

interviewr

Yeah. So now we, we'll start with data. Um, did you ever use the, the services of, um, data collectors? So people that manually. Create data for, to train your model.

interviewee

Yeah. To be quite frank with you, I, I think something like, uh, some sources that we were using, uh, they were manually, uh, inputted, um, Yeah, like, I guess these reports like that, you know, like they would take an analyst and then someone needs to put it into the system, and then the system would bring.

Data-a-lot-of-manual-work	Data-collection-manual	Data-unstable-provider

interviewee

To us, like say to the next level so we can access it. But I'm, I'm, I'm pretty sure that yes, manual, manual inputs were there and we would see it because sometimes, like one week, sometimes it would make no sense, some values, but overall it made sense. But some, some, some transactions, let's say some, some roles on your, on your table like would make no sense.

Data-a-lot-of-manual-work	Data-collection-manual	Data-unstable-provider

interviewr

Okay, I see. And how, how can you enter the quality of the data when, when there is issues that.

interviewee

Um, so I guess if you have, I guess you, you try to answer to these questions in the exploration phase, right? Like, when you're trying to decide, like you have a, you have a, an ML problem that you, you are trying to solve, you think of the different data sources that are gonna be useful, then you need to go see the data and then you need to evaluate how reliable the data is, right?

interviewee

Like, you know, is it changing on a weekly basis, for example? If I pull the API for last week, is it gonna gimme the same number this week, the next week, then the week after, or not? You have to know these things. Um, so I guess for me it goes really in the evaluation phase. Like when you're doing the exploration, it not only the exploration of, uh, yes the data is correlated to my, to my actual target, but yes, the data that I'm getting from this API is consistent.

interviewee

It's always giving me a range between this and this. And. You know, like I cannot expect a different type of, like different format for my data. Like I can always expect the same columns and stuff like that. So I guess, does that answer your question? Like, for me it's in the, uh, exploration and then data source evaluation.

interviewr

Yes, totally. Thank you. Um, so did you ever use external data such as public data set, uh, third party API or web script data?

interviewee

Yeah, all of the above.

interviewr

Okay. And did you ever add qualities with the data?

interviewee

Um, Yes. Yeah, Yeah. Okay. Like, like another example before, uh, like for example, that's, uh, that's one, right? Like, it's a, it's an api.

interviewee

We had quality issues on that one. Uh, on the government data, it was more the fact that it was not running, like the API was not available. Um, and then scraping. Not quality issues of the data itself, but like more about like the website changing and then you have to like, change the scraper. Um, Okay. Yeah.

Data-unstable-provider

interviewee

But I, I guess like something, you know, like something with it, like we did also like, it just sticks in parameters. Like, you know, for example, like if, if you're, if you're getting something like, uh, like number, like if you, you're expecting, uh, whatever a number, you know, the range or you can have an idea of the range that you can.

interviewee

Like, if you have negative values, for example, like, you know, that, that doesn't make any sense. Um, you ignore those, uh, or you flag them again, for us, what what we try to do is like just to flag as much things as possible. Like it requires human effort after the fact. Like to go and see what, what's going on.

interviewee

But at least like, you know, you, you keep trustability. Like if, if something occurs, you can go back and then see what. I see. I dunno if I answer the question.

interviewr

Yeah, yeah. Thank you. Uh, did you ever use data that was generated by another system? Uh, it can be ML based or not by another

system.

interviewee

Can you, if, if, if it's not an ML system, like what, what would it be?

interviewee

Can you gimme an example or,

interviewr

Yeah, sure. Um, so you could have a model that, uh, use that uses by sale to predict the future sales of a project, and the by sales are generated by a system like transaction, you know, uh, or it can be sensors, for example. Yeah.

interviewee

Yes, yes. For, Yeah, yeah, yeah. So sales for example.

interviewee

Yeah, I use that. And, um, and yeah, like there's, there's somewhere, There's some crappy stuff in that. So that's what we call the internal data because we were using that and, uh, you know, the, um, for that we basically use filters. Um, filters. Yeah, filters in terms of, you know, for example, sales from customers.

interviewee

Like, I only want the customers that, in my list of customers that I, that I. Like, don't, don't gimme crap that, that I'm, that I'm not interested in. For example, the values of the transactions, right? Like if you gimme a transaction that, you know, if I'm expecting value between SIR and 20, like if you gimme a hundred, like I know that doesn't make any sense, so ignore those values.

interviewee

Um, so yeah. But basically we, we, we, we protect by using filter.

interviewee

And then filter out. Yeah, go ahead. I mean, yeah. I have a follow up question. And by filters basically you are targeting some, ignoring the data. Yes. But you are explaining that vendor then there

interviewr

is a problem.

interviewee

Simple approach or approach that you are using is just ignoring those data. Yeah. Perfect. Thank you.

interviewee

Some cause, some, sometimes, but for that you need, you need the, the business. I guess like you need the business, uh, input, right? Like for example, if I, if I don't know anything about the business, like it's, we need to find out like, you know, having a hundred here, is that real or that's an error? Cause like, if it's real, then like, it's an outlier, right?

interviewee

Like, and that's a different question. You need to, uh, you need to build your system. So like, it's either good or bad that are predicting outliers, or you have, you have to have an idea of how to use those outliers, how it affect the. But if it, if it, if it, if you show it to the business and the business tells you like, Yeah, that never actually happened, that's a mistake that someone typed in, Uh, that's when you just ignore them.

interviewee

So I guess like it really, it really depends on, um, on, on, on use case. By use case. Does that, does that answer your question? Yeah. But if go. You are saying that some domain knowledge or knowledge about the business, you have to know Exactly. Yeah. So domain, domain knowledge and I guess company specific knowledge, right?

interviewee

Like maybe you work in the same domain in different company, then you would not know where like that was the case. But yeah, like you can it as a as domain knowledge. Yeah. I see. Thank you. Go ahead.

interviewr

Thank you. Uh, I can ever measure the quality of your data and or try to improve

interviewee

it. Have I measured the quality of my data?

interviewr

Yeah. So, so you mentioned you had quality shoes. Um, I, I guess you, you look for quality shoots in your

data

interviewee

and have I, So, so that's so, yeah.

interviewr

Uh, basically I will, I will ask you. So, so the follow up question is, do you have any tools or framework to help

you clean or transform your data?

interviewee

Um, honestly, nothing, nothing really automated.

interviewee

Like, it's never a put data in get me data out that it's clean. Like, it's never like that. Like there's always an

exploration phase where you, uh, you know, you, you, you use visual, visual. Visual tools, like, you know,

plotting, I don't know, like the distribution or, you know, like if it's a time series, you try to correlate, you try to

do a bunch of stuff, but basically you try to visualize the data so you can find out things that are weird or

abnormal or, Yeah, basically abnormal.

interviewee

And then once you identify those, like it's, it's really knowing like, okay, why, like, is this actually gonna happen

on an ongoing basis? Like, is this. Is this more like, most likely a mistake that like a person typed in by

mistake and doesn't make any sense? Uh, but so yeah, it's, it's, it's visual. It's visual exploration.

interviewee

Like never, never, like I use this tool and like I do like dot clean data and then does all the magic for me.

Never, I've never done it like that. Cause I, I think, I think very, like, it depends on a, on a case by case basis.

Like I don't think there's, there's one solution that would fit all problem. Does that answer my question?

interviewr

Yeah. Perfect. Thank you. Uh, is there any other data quality should be missed that you consider

interviewee

relevant

interviewee

data, Data quality? Um, so data quality. More, more stuff. Let me think. Let me.

interviewee

Not, not quality as per se, the quality being bad, but like, let's say if something happens in the distribution of your data, right? Like let's say you train the model, uh, I don't know, for example, Covid, like you train a model on Covid and then like you hit Covid hits and then you keep making predictions like, like you're actually underlying process that is generating your data.

MM-drift

interviewee

That changed. So it's not, I would not say it's a quality issue, but you know, things can happen in the market. For example, whatever that is, changing the process, that is generating your data and sometimes like, you know, like you miss it. Like if you have something running, you miss it and then you're screwed.

MM-drift

interviewee

Like, you can make, make predictions assuming that you have the old processes. So it's not, I would not say it's a data issue, but that's important. Like you, the data drift, like. Like if it's changing, like having some parameters for your data to analyze your data to break it down, Say like, Okay, this is consistent with what I've seen before.

MM-drift

interviewee

Does that, does that make sense?

MM-drift

interviewr

Yes, it's a good point. Yeah. Totally. Totally. All right. Uh, so we talk a lot about data now. Now I'll move on to, uh, model. So my first question is, how do you evaluate the quality of your models?

interviewee

Um, so having, so I guess like in this case, I'm gonna, Speak more about, I mean, I, I can, I guess I can build something general, but like, of course in the, in how you were comparing the performance on, on your train validation test, um, and then across different things, like for example, if it's a forecasting problem and then you have different personalities, you wanna make sure you wanna understand how good your model is across time.

interviewee

Right? And sometimes that's an issue because you don't have enough data. But ideally, like with enough data, you would know, okay, like I wanna, I wanna look at how my model does across time. Cuz like, maybe your model is really good for the summer, but then you're gonna be screwed for the winter, for example.

Data-low-quantity

interviewee

And I guess if you were doing a classification to really make sure that, you know, like you can slice and dice your data in different ways in, in then make buckets of, you know, like different things that you're classifying and then see like, okay. , you know, like, am I classifying as good when I have this variable?

interviewee

Then when I have another variable, like is there something specific to the data that I'm not being able to classify? I guess you could also do that in classification. Um, but I guess overall it, it sums up as, you know, like analyzing the performance of your train and then your validation and then seeing that.

interviewee

That's consistent. Um, and then of course once you deploy that, you keep having that what you're expecting, right? Cause like, I guess like you had your relation and you have your test set and that, that's what tells you like, Okay, I'm, I'm good, I'm good to go for my model. And like that once you go in production, that you're getting something similar to what you were expect.

interviewee

Does that make sense?

interviewr

Perfect. Yes, it's clear. Thank you. Um, have you, have you ever used benchmark model to evaluate the quality of a model you

interviewee

are, uh, building? Yeah, so we, we always try to do a, like a simple model. Uh, sometimes that benchmark can be like, let's say forecasting can be, um, you know, just an adjust and uh, a moving outreach or it can be like, same thing last year for example.

interviewee

That could be like a benchmark. And um, you know, like if it's a classification, like if it's a binary classification, like for me, the, the benchmark is always, you know, what's the ratio of my, my true versus false, right? Like if you have, for example, 90% of observations are false and then 10% are are true.

interviewee

Like my baseline is 90%, right? Like my accuracy 90%. So if I have a, if I have something, uh, that predicts to false, everything. I'm gonna have an accuracy of 90%. And then you can evaluate your other metrics, like precision recall, whatever. So that could be like, for me, a benchmark, let's say when doing binary classification.

interviewee

Um, and then like, you know, like you have other types classification, but like multi-class. But for me it's, it's that. And then, um, Yeah. Does, does that answer your question? Yes, it does. Thank you. But always, yeah, always, always. That's a rule of thumb, like always, always have a benchmark. And it's better when that benchmark comes from the client.

interviewee

Like when it's better when they have something that they're already using and then you're trying to improve. Like that's, that's the best that I think

interviewee

you're sounds the quality. No, you're good. I can, yeah.

interviewr

Okay. Okay. Have you ever assessed the quality of ML model prediction using this, using the users of this system? Say

interviewee

it again,

interviewr

sorry. Uh, have you ever used, have you ever assessed the quality of ML system? Uh, with its, with its users.

interviewee

With the users. Um, Yeah.

interviewee

Yeah. And I guess like that's those, that's one of the toughest ones cuz like, don't, like sometimes, you know, like you can have an ML system that like, it really does, it's, it's fully automated, like the predictions or the output of the model. It's the input for something else and then that's something else makes a decision and then something happens.

interviewee

But like the issue is when it's a user like that needs to make a decision based on what the model is. , That's always the trickiest part. Like it's the most difficult part, I think because maybe your prediction is good or somewhat good, and then the, the user is not able to make the relationship of like, Okay, now that I have this prediction, how do I apply it to my decision making process?

interviewee

Right? Like, I have all these different things. I have this model that, you know, I know it's technology, but they don't really understand. And then so they. If they don't underst strength, that if they don't understand the strengths and weaknesses of the, of the, of the system, then they don't trust it. Or even if they trust it, they don't know how to like, take that output from the, from the system and then make a decision based on that.

interviewee

So, I, I'm answer your question, but like, for me, this is the most, is the most difficult thing. Like, so, you know, it's a, sometimes they, I guess users expect a perfect predict. And it's really hard, almost impossible to have. Right? Like they, they expect the system to be a hundred percent perfect, and if it's not perfect, they don't, they don't want it.

interviewee

Which makes, you know, like it's a, it's a wish, wish in there. It's like, it's, it's a mistake, I think. Not a mistake, but, you know, like it's a, it's a problem.

interviewr

I see. Thank you. Um, have you encountered any other quality issues during the evaluation of

interviewee

your model?

interviewee

Uh, quality issues.

interviewee

Can you gimme an example or you're supposed to Not supposed to.

interviewr

No. I mean, you, you mentioned some. If you don't have any other idea, that's fine. We, we can, Other issues

interviewee

with my, with the model you said, right?

interviewr

Yeah. That you encountered during the model evaluation phase.

interviewee

Um, so the model evaluation.

QM-explainability-issues

interviewee

Um, No, sometimes it's tricky, uh, to get the, uh, feature, like, like what's, what's causing a prediction? Like the feature attribution if you want, or feature importance like sharply values, stuff like that. Ah, you are talking about explanation. Yeah. That's, that's, so you, you trying to explain the. Yeah, that's, that's, that's very, uh, in my experience that it's always very tough.

QM-explainability-issues

interviewee

Um, and did you find it

interviewr

helpful if you do

interviewee

such evaluations? Say it again. Did, did you find it helpful,

interviewr

for example,

interviewee

doing explanation study, trying to understand what effective model decision? Is it helpful? So right now, right now, for a project I'm working on, we have a tool that's supposed to be for the user to play with the features.

QM-explainability-issues

interviewee

And then to infer to, to make an inference. And you know, like just, just by using it ourselves as similarly, we were the user. Like we realized that it's super useful for us as data scientist. We really understand like how, like, how the model is using that. And then, you know, like making prediction and what is the user expectation.

QM-explainability-issues

interviewee

Exactly, so, so, so that, yes, for the user, but also for, for, for us, like let's say if I'm a data scientist, I'm thinking that some feature has a positive or negative impact and I test it and it's not the case, then, you know, like maybe my model is not learning the right thing. Like maybe I'm, I'm missing something.

QM-explainability-issues

interviewee

So like, so it, it's, that's super useful, I think like just being able to, uh, to play with the features and then see, But with the visual tool, right? Like I, I know I can, I can change some stuff in. My ex predict, and then you just inference, inference season numbers. But when you have a visual thing that you can drag the values of your feature and then see how that's changing your, your prediction.

QM-explainability-issues

interviewee

That's, for me, that's the most useful thing. And then we're, we're like, just by mis, not by mistake, but like, you know, we have, um, again, we're building this for a, for a user, an end user that is non-technical and we're noticing that for us is super useful as data scientists. Perfect. Thank you. Yeah.

QM-explainability-issues

interviewr

All right.

interviewr

Uh, so great. We will move on to the deployment of machine learning software system. Uh, so all and where are your model? Deploy.

interviewee

So from what I've done always in the cloud, like always for like a third party provider that you know, is hosting either a virtual machine and then you have all your logic there.

interviewee

Um, and you know, like it's, it's basically reading the data and then doing the inference like on a schedule basis. Or with APIs, uh, but it's always hosted by third party, uh, third party providers like Microsoft or aws, uh, Google.

interviewr

I see, I see. And what are the challenges we have encountered when deploying machine learning software system?

interviewee

Um, so I don't, I don't think I'm a super expert in, in, in the deployment, but you know, what I'm seeing is basically not having an estimate. How, um, the, I guess, like, what's the right word? Like, the intensity with which the model will be used, right? Like if you have, if you, if you, if you have deployed in a small machine, like, and then, you know, like all of a sudden you have like a thousand requests that come in at the same time, then, you know, like it's, it's gonna, it's gonna block, it's gonna explode.

QM-scalability-issues

interviewee

Or like, you're not gonna be able to serve your. Your, your clients. So I guess that's something that I've seen, like not having, not knowing what the true demand of your model is gonna be. You are referring to scalability. Scalability, Yeah. Like it's, of course, it's like, you can put it like to scale infinitely, but like you have more cost associated with it, right?

QM-scalability-issues

interviewee

So like you always try to like go with the cheapest, with the smallest, uh, but then sometimes the smallest is not enough. I see. But yeah, with scalability, like if you have, if you have money to pay, like something that's self scalable, like, then I guess you're, you're good to go, right? In theory. Mm-hmm.

QM-scalability-issues

interviewee

thanks. Um, yeah, that's, that's, that's the only thing that I can come up with.

interviewr

Perfect. Uh, so moving on to next question. Did you ever have a model that performed well locally, but when you deployed it performed bad?

interviewee

No.

interviewr

Okay. And have you encountered any other quality issue with your model or system during deployment phase?

interviewee

I've encountered Anything else? Um, like it depends on the, on, on how you're, So a lot has to do with how you're set up cuz you know, like, usually like we develop locally, And then once everything is ready, like we, uh, we deploy or we're, we're test, Yeah, we deploy, I guess like to a, to a development environment.

interviewee

And you start seeing how it works in the cloud. If you're not properly set up, you can run into like issues, like dependency issues, like, Oh, I, I'm not using the right python, or I'm using like another, basically like, you know, like a, an old version of a giving software or a new version. So that's, that can be opinion in the as if you're not properly set up.

MD-issues-with-dependencies

interviewee

And you're not, uh, sharing all your, I guess if your development environment is not unified across the different members of the team, then it becomes a problem because someone's developing with one Python, another one's developing another Python, and then when you try to like, put everything together, like then it explodes.

MD-issues-with-dependencies

interviewee

So, um, so that's, that's an issue. But again, you solve it by having a unified development environment. So right now we, we have. For example, we, we have everything containerized. We have a container where we have all the dependencies. Everything is unified across all, like the engineers, data scientists. So like that, like whenever we push, like we can deploy automatically and then everything is, everything is unified.

MD-issues-with-dependencies

interviewee

All versions. Yeah. Okay.

interviewr

Does that answer your question? Yeah, totally. All right. Um, Some way onto model maintenance, Uh, actually maintenance of machine learning software system in general. Yeah. All the ensure that its quality doesn't decrease over time.

interviewee

Uh, so for me, I'm, you, I'm a, I'm a big pusher of retraining.

interviewee

But if, I don't know, if, like, again, like I don't, I don't, I don't think I can ensure that the performance will decrease, will not decrease over time. Um, I wish I had something like that, but you know, like for me it's, again, like it's more on the, uh, when you're doing your training, like, and then you're evaluating your, your, your performance.

Issues with performance alerts   MM-drift

interviewee

That's, that's when you try to think that, you know, it's not gonna go. Your, that your performance is not gonna go away. I guess you focus on that there. Um, and then if you can have like a deployment phase, like, which is like soft deployment if you will, like where you start testing it. Um, but you know, like in retraining, like for me, I'm retraining.

Issues with performance alerts   MM-drift

interviewee

But again, like even if you're retrained often, then you know, like if something happens, like again, like if something happens in the underlying processes that generate your data, Doesn't matter if you're retraining, like you're still gonna be looking at the past. Um, I guess you can, you know, there's different parameters that you can use, like, I guess retraining and then only looking at more recent data, focusing on, on, on more recent data.

Issues with performance alerts   MM-drift

interviewee

But even that like doesn't guarantee that, you know, performance is gonna be a good, I guess the, the best is like to know. So once you know that how it performed in the past on a validation set and how it's been performing, I guess the best would be to have a notion of, you know, our, the, um, it's like a confidence interval if you will.

Issues with performance alerts   MM-drift

interviewee

Like how, how confident my, my system is in making this prediction right now. I guess that would be the best, like, cuz you can really ensure, like, it's more about like how confident I that of the prediction I'm making. Right. Yeah. That's, that's all I have.

Issues with performance alerts   MM-drift

interviewr

That's, that's a good idea. Yeah. Um, actually have very encountered issue with data during the model maintenance.

interviewr

And it can be the data itself or the, the data source, which could be unreliable,

interviewee

for example. Yeah, yeah. Like, so what I explained before about the weather data, that was a big one.

Unexpected. Um, cuz it, it, it was unexpected because it changed, Let's say you get weathered for, you know,

like even if we got weather for.

interviewee

I don't know, like months ago you wouldn't expect the weather to change, right? Like, like the weather. Like

okay, if I tell you like, what was the weather last December? You gimme a number and if I poke you one like

six months later, and if I tell you what was the weather last December, Give you a different number, that was a

bit unexpected that so far in the past I could change something to something that made no sense.

interviewee

So yes, I encountered that. Yeah, that's right.

interviewr

That was a good story. Yeah. Thank you. Um, yeah. Have you encountered issues with the model during the

maintenance of ML system? I, I guess

interviewee

yes. Purely with the model. Hard to say. Like I would say no. Like it's always data related, right? Like I don't

think it's a model for that.

interviewee

We try to build models that are not super, that overfit and that, that are not super complex. Um, I've had bad

experiences in the past, uh, building models that are super complex, super good in train, maybe valid, but like

once you go in pro, like something, some little something, little changes and then. It goes all over the place.

interviewee

So for that, we try to use models that are less complex, um, you know, less, less features. Like, again, it

depends on the problem that you're trying to solve, right? But like, if, if you can, like, I would like, for me, I'm

always like less features. The better, the less complex the model, the better. I'd rather have model that is less

good overall, a bit less good, but it's more robust than having something that's super good.

interviewee

But you know, like something little's gonna fuck it up. Did I answer your question?

interviewee

Uh, question. Okay. Now I can hear you. Were you on mute for a little bit.

interviewr

Okay. That's okay. Perfect. Um, so did you, do you have any other issue regarding the maintenance of

national learning

interviewee

software system? So the issue with the cloud, like just the, the, the, you know, like the triggers don't work.
Um, we've had issues also with the email.

Issues with performance alerts

interviewee

Email, so like we have. We have email alerts whenever something doesn't run. For example, like we have a
pipeline running that's supposed to run every week. If something doesn't, doesn't, doesn't happen, like
something fails, we have an email alert. Sometimes those email, email alerts fail because of the underlying,
you know, like, um, services that we're using.

Issues with performance alerts

interviewee

Sometimes that happens. So we don't get the alert and you, so you don't even know that it failed. Um, but
then, you know, that can be solved by using something more robust. And the cloud, like the cloud just, you
know, like your, your pipeline not running for reasons that you don't know. Like it just stays spot as I explained
before.

Issues with performance alerts

interviewee

Um, that's pretty much it. Yeah. Perfect. Thank you.

Issues with performance alerts

interviewr

Uh, alright, so, so now it's the third question. Um, the. Sorry, there is three more question, and, uh, basically I'm gonna ask you if you ever had issue with any of the following quality aspect of machine learning. Uh, and if you don't, there's no problem. We are just probing.

interviewr

Uh, so did you ever add issue with fairness?

interviewee

With what? Sorry?

interviewr

Fairness. Fairness. So furnace,

interviewee

it's not like, You mean like ethics related? Yeah, like for like, you know, like I'm not discriminating in some, some type of people. I've never, not essentially people even, for example, type of data,

QM-fairness-issues

interviewee

if you feel that the model doesn't perform fairly according to the, That's, so if, if you see, Repeat that again please. I. Uh, anything, any, any, any moment that you feel that the model doesn't perform fairly, uh, regarding the inputs? It can be a people, for example, making, making different decision against women or men.

QM-fairness-issues

interviewee

It can be, Yeah. So I, I've never, I haven't, um, I haven't encountered that and honestly, it's some, it's not something that, Maybe because I've never had a problem that was so where furnace was

QM-fairness-issues

interviewr

so important. That is,

interviewee

that is the case. It's problem, It's a domain dependent. It It's domain, yeah. Like for example, I worked at a bank, and then we would not look at that data, right?

QM-fairness-issues

interviewee

Like for example, sex or stuff like that. Like we were not, like we wouldn't, we wouldn't touch it. Like we wouldn't even look at it because we know that it would bring, it would potentially bring some predictive power to, so we were not even allowed to use it. Um mm-hmm.

QM-fairness-issues

interviewr

Because they, they, they might impose some

interviewee

bonus issues.

interviewee

Exactly. Yeah. Yeah. And

interviewr

in that case, have you verified that your model was still fair, even if you didn't use, um, the sensitive features?

interviewee

Good point. No, I did not. Um, I did not. I did not, but I didn't have, I didn't even have that data to test it. Uh, a good point, but I did not. And it's honestly something that.

interviewee

I don't like, you know, I make fun of my friend because she's very ethics, She's also in AI and she's very ethics, uh, minded. And I always make fun of her because I'm like, uh, who cares? Like, as long as you're predicting good, like, but you know, like it's, it's a, it's a fair point. Like I think we're gonna see more of that in the future to evaluate friends of, uh, of models.

interviewr

Okay. And I just wanna clarify, clarify something you said. So you, you said you cannot, uh, test that your model was fair. Is it because you didn't, you didn't have access to the sensitive feature? Exactly.

interviewee

Yes. Yeah.

interviewr

All right. Uh, did you, uh, ever add issues with robustness with your model?

interviewee

Uh, yes. Yeah. For, Yeah.

interviewee

Yeah. So for example, like that, And what were they? Yeah. It was a model that was too complex, a model that needed to be super accurate. It needed to be super accurate. Like that was like the, the, its purpose dependent on need being super accurate. And for that we build a model that was very complex and I mean very complex, you know, like it was a deep learning, but you know, like I'm sure like Google, like translation, translating systems are, are, are bigger.

interviewee

But like, you know, for, for my use case it was a, it was a complex model. Um, and then once we put it in production, like it worked well for a bit of time, then something like some disturbance is happening in the market and then. , you know, like it was not, it was not nearly as accurate as what we had seen before.

interviewee

Um, and yeah, so.

interviewr

Perfect. Thanks. Uh, did you ever add issue with expandability? Yes.

interviewee

That's, that's one of our main issues. I would, I would think, yeah. Um, ongoing, ongoing issue. We, we use a shapley. Um, like as, I guess if you, the framework that we use, like to try to like, build explainability, but even that like, you know, it's not, it's not a one size fits all.

QM-explainability-issues

interviewee

Like you cannot just do where you know, or get ly values and then aggregate those and it's, it doesn't really tell you the full story. It's just one way of distributing your, like, you know, like, I guess like the, uh, the weight or the effect that your features have on your prediction. doesn't mean it's good, right?

QM-explainability-issues

interviewee

Like if, if you have a crappy model with a bunch of features, like it's just gonna give you some random stuff. So that's, that's, that's an ongoing issue I would say, like, that's, that's super important. Like for me personally, and it's an ongoing issue, I would say we haven't, we haven't found the recipe like it, I guess.

QM-explainability-issues

interviewee

I guess a lot of it, it's like domain related. Like you have to look at the features, see what the feature, uh, attribution is telling you. See if it makes sense from a logic business point of view. Try to validate like that. And then like once you have something that works more or less good, then you are like, you deploy that, you know, that, how can I say, like that, that feature importance, feature ion system.

QM-explainability-issues

interviewee

Uh, but yeah, it's, it's an ongoing pain. Yes, I would say .

interviewr

Yeah. I see, I see. Thank you. And uh, finally, do you have any issues with the privacy,

interviewee

privacy?

interviewr

Yeah. Could you be model, could leak am model that could leak? You have to be careful. I will give you one example. Uh, for example, if you have a complete system and you know, when you write emails with Gmail, sometime the, the searches you some, the next word, well, potentially, if it is trained on a lot of data that can be sensitive.

interviewr

It could leak. Uh, if you al always choose an next word, maybe you, you would leak private information from another email, you know? Uh, so this is one example in the general concept, it's just that it leaks the model leaks data and private information, uh, somehow. Yeah.

interviewee

Uh, no,

interviewee

no. But also, like the models that I've done are always for internal use. Like they're not for, uh, Like from a company to end users. Like they're always like for the internal use of the company, for example. So like, it's, it's, it's, it's internal data. If it's leak, like, you know, it's, it's internal. So it hasn't been a mm-hmm.

interviewee

a concern.

interviewr

I see. It's not so much problem. Yeah. You know. All right. Uh, so pretty much the last question. Um, in your opinion, what is the most pressing quality issue, uh, researchers should try to solve?

interviewee

So I would say explainability is a big one. And robustness, like for me, those are the big thing. Two things, maybe robustness more than explainability, but robustness for me, it's a big one.

interviewr

All right, thanks. And do you have any other comment about the quality of ml?

interviewee

No, I just hope I can keep learning. Uh, I'm, I'm, I, I, I, I mean, I'm by no means, I mean, it's pretty new what we're doing, right? So like, I guess like anyone with a bit of experience, it's an expert. But you know, like for me, like I have so many things to learn, um, and I'm always learning. So like, I hope I can keep learning so I can, uh, know more and maybe contribute something to.

interviewee

To the world of ml. Right.

interviewr

That's great. Final thoughts. Right. Well, well, anyway, thanks a loto. Uh, really great to have you here. I think it's gonna be really interesting for our project. Uh, so, Okay.

interviewee

Yeah. Thank you so much. It was very interesting. I hope. I hope it was useful. No, for sure it,

interviewr

for sure. Thank you, son.

interviewr

All right, cool. All right. Bye, Bill. Have a good day. Bye.