

Interview 25 - PO

Interviewer 1

All right. Uh, to start off, can you give us some information about yourself? Uh, so how many years of experience do you have in machine learning or in general, anything like

Interviewee

that? Yeah, yeah. Uh, my name is X. I am currently, uh, machine learning engineer, uh, in, uh, company here in Country X. Uh, I have roughly about, uh, five years of experience.

Interviewee

I started as a data analyst, uh, with a corporate company, and then switched more, uh, into, uh, startups. And I have worked first with a startup in, uh, uh, something like, uh, it's kind of FinTech and e-commerce together. Uh, if you are familiar with Groupon, uh, it's something like that started as a Groupon and with the AI its became kind of a FinTech.

Interviewee

I was the first data scientist there and, uh, that's where I have been, uh, basically exposed to work on projects of, uh, kind of a like full stack or end-to-end nature of our project. And, uh, later on I moved on to, uh, another company where, uh, it is still a tech company, uh, that has a platform, uh, and both business c2 B and B2C for cars and or secondhand cars, basically, uh, ecosystem.

Interviewee

And, uh, I led a small squad of data scientists there and, uh, later, last year, It kind of feel weird to say last year because it was just a week ago, but , uh, yeah. At the end of last year, I, uh, switched more into, uh, basically focusing on, uh, ML in production for, for the whole, uh, data team. Uh, I would say most of the, uh, problems or use cases that I worked with are, uh, along the lines of recommendation engines, uh, prediction engines, uh, like regression and so on.

Interviewee

And, uh, briefly I touched on use cases, uh, in, uh, uh, search engines or semantic search engines. Uh, like that.

Interviewer 1

Yeah. Okay, great. So we're glad to have you. Uh, so I'll start off with the first question. What are the main quality issues you have, you have encountered with your data model or system so far?

Interviewee

Uh, I think the most, or the, the first problem that, uh, pops to my head is, uh, uh, data quality issues.

Interviewee

Uh, uh, for example, like, uh, eventually when, when there is, uh, unexpected data type that comes in, uh, that is rhinos because of upstream, uh, basically fault and upstream validation of the data. So that kind of, uh, data quality issues is the number one enemy, basically because, uh, often that is not, or I don't know how it'll vary, but in my experience and in the, in, in the systems that I have, uh, it is something that happens sometimes and, uh, it disrupt the whole thing.

Interviewee

So like, uh, uh, especially if you, if you consider the use case that I have where there are auctions about. Take place, uh, about four times a day. And the nature of the recommendation, uh, solution that we have is because our inventory changes for each auction session, and therefore our basically model needs to be retrained based on the inventory of that session.

Interviewee

So if it fails on that time, then uh, it's a very short time to, to look up and figure out exactly what is this error or what is happening. And it's, uh, the most disappointing when it's like, oh, I did not expect this data type to, to appear, assuming everything is integer and that somehow some string found itself.

Interviewee

Uh, something like that. So this is number one, uh, uh, I think issue. Uh, and uh, I'm not sure like to what extent, uh, other companies will have them. Uh, but number two I think is. Uh, generally just observability. Uh, so often if there is, uh, let, let's say certain teams that are using a prediction model and they started questioning the predictions or whatever that we are giving to them, considering the use case that we have is, uh, what you can, uh, call it as a decision support system.

Interviewee

So like it supports these people to take the decision. It's not, uh, something that they have to follow a hundred percent. Uh, and uh, so like if they start questioning, uh, the, the predictions that we have or something, then, uh, things along the side of when, what model is this? When was it trained on? What data was it trained?

Interviewee

What was like, perhaps the, the timeline of trained tests split, for example. All of these things. Yes, they are there somewhere, but they are not easily retrievable. The, so like, you know, for someone to go and look into these things, it, it, it'll be like taking some time before, you know, we are able to, to touch or have a direction into, uh, the right issue that results in a, a questionable, uh, predictions.

Interviewee

Uh, these two things are, uh, the main issues that, uh, I see. Uh, there is another issue that we have, uh, which is more, I'm not sure how it, it would be related, but if you would consider the quality is like, uh, perhaps like the faith of the people who use this as part of the quality certain times, especially when we would work with.

Interviewee

Uh, imaging problems? Uh, uh, perhaps, uh, one of the things that we would see often is that, uh, teams do not, uh, perhaps they do not, uh, see, uh, so for example, like there is 30,000 images that we predict, for example, on, or we do some annotations on, and there is like few of them that has some error. And, uh, in, in a way, uh, uh, basically it's not easy like to figure out why, uh, why why is basically, uh, is this, uh, uh, mis annotation is taking place.

Interviewee

Is this because of this image or the nature or in the environment or the orientation in which this image are taken is not represented into the, the data set. Or how, for example, to communicate that this is, this is not necessarily like a, a, a quality issue because we have a certain accuracy level that we maintain.

Interviewee

Um, yeah. But it's more of like, um, I think, uh, I, yeah. Uh, , uh, the, the, the, that, those all the things that comes to my head.

Interviewer 1

Yeah. Thank you. So for the last issue you mentioned is a problem that you have wrong annotation or that the model makes wrong prediction?

Interviewee

Uh, the mo So it's mostly like the model will be doing some form of, uh, object detection.

Interviewee

And when the model doesn't do that, for example, uh, it, it'll be raised. Investigating why, for example, it becomes very hard and, uh, uh, perhaps because maybe even if we have data sets, especially like with imaging data sets, that is, for example, for object detection, we look at them as image, and we just care about the boundaries and, and all of that.

Interviewee

And when, and a problem happen and we want an interpretation, there is no, no such thing. Like we cannot say, for example, uh, glaring image. And we have this percentage of the image that comes with glares, uh, such, uh, x image that you can call them, perhaps like awkward angle image or something like that.

Interviewee

Like in your training set. So when, when some misclassification happened, or I don't wanna call it misclassification, uh, let's say mis uh, annotation happened from the model side, you kind of, you can pinpoint like what, what might be the issue. It, uh, I think that's kind of, uh, contributes to the general quality that, that, uh, you know, such things.

Interviewee

I don't know, perhaps maybe there could be things put in place early on that makes it easier or, uh, yeah. Um, uh, yeah, I'm writing out on how is, but No, no, it's perfect.

Interviewer 1

Thank you. And so what I understand is you have some prediction that are made up by your model, by a model that are wrong. And what you're trying to do is to understand why, why the model made or wrong prediction.

Interviewer 1

Um, and how do you achieve that? Is it by looking at the picture and guessing, or do you have some tools to help you with understanding the model?

Interviewee

Yeah, I think the absence of these tools or practices on place is the, the main thing. Because often when there is something like that, just someone has to look at it and start.

Interviewee

Brainstorming, I guess, to come up with a reason that is justifiable to make people basically say, oh yeah, okay. That can be a reason. Otherwise, it's, it's just an image and we see that fault, but we can't say why. Okay.

Interviewer 1

I see. Thank you. Uh, you also mentioned, um, data problem and or observability issues. Right.

Interviewer 1

Um, do you have any, did you put any mechanism in place to like atin weight your issue? So for example, for data, you said sometime it's another, uh, data type instead of a string, it's an integr. Uh, do you have something put in place to detect these kind of problem and fix them? Maybe automatically?

Interviewee

Uh, okay.

Interviewee

Yeah. Uh, but usually, U usually the, the, the way it happened is that, for example, like you, uh, you would look at the last six months or one year and you would train your data on, on, on something like this. Uh, but then, uh, uh, when, when you are deploying, uh, the solution, for example, the, the, the data looks fine to you, especially if you are converting between, let's say, numeric type of, of, of data.

Interviewee

Like there might not be a bigger issue, but the issue comes, uh, if, uh, there is a fault up upstream that contributes to this, uh, uh, to this maybe I would call it sabotage of expectation. So that you're expecting things to be a numeric and you kind of. Cast them from one data type to another, from into float or vice versa.

Interviewee

But the thing that came from upstream is something completely different. Uh, where after it happens, of course, like you figure it out and you're like, oh, okay, maybe we need to do some level of filtering, for example. Uh, but, uh, as far as quality of ml uh, systems, uh, this is one of the issues that, uh, of course, uh, contribute to, uh, basically harm or jeopardize the, the quality of things.

Interviewer 1

Okay, perfect. Thank you. And for the second point of observability, you mentioned you have some, um, system to um, uh, see well, to have some observability in your. Sorry, I'm, I will rephrase that. So you mentioned you had observ observability issues, and I guess you have some tool to, um, understand your system.

Interviewer 1

Uh, can you tell us what are these tools?

Interviewee

Uh, I don't necessarily have tools just to be understood. Yeah. Uh, so, so for me, I'm the guy of who will always just keep, keep data, but keep versions of the data store, the ministry and keep them there. Uh, uh, but, uh, the more I see how, how severe this and questions will come, like, oh, we have this, uh, mis predictions and, uh, what is the model that is live?

Interviewee

And questions of that thought. It, it start, it start makes me feel. This is bad. It needs something to be done. Uh, I have br tried to bring stuff into place, but this is a general practice, basically. Uh, so one of the things I tried to bring was, uh, D V C or data version control. Uh, and to be honest, me, myself, I couldn't stick to it and let alone the team.

Interviewee

Uh, it, it, it, they weren't that much of fans, but I was doing that in my capacity as a, as a, as data scientist or a, as a lead data scientist, not where I am today as ML ops, because ops is more about setting the tone, I think, but the way it's practiced in the organization, I work. Uh, another thing that I try to bring and it start to gain traction is, uh, ML flow.

Interviewee

So I brought ML Flow to kind of say like, Hey, look, you can train thousands of models or anything, and we can know what version that you promoted to production. So if someone say, oh X or Y model, we can retrieve it and we can know what data has been trained on and, uh, stuff like that. Uh, I wouldn't say it's like fully adopted and accepted, but it is gaining traction.

Interviewee

Few people in the team started using the ML flow deployment that, uh, intuitive to push. Yeah. Uh, mostly the, these two things, but uh, in regard to the first point about the quality of the data, uh, uh, for. Reasons I might not like, uh, bring here Just that, uh, yeah, the, the leadership doesn't see, or doesn't necessarily, uh, find this to be very, perhaps a pressing issue.

Interviewee

Not, uh, it doesn't find a pressing issue. Let's say they see the solution in, in a different way. So for me, I would see a solution in things along the side of, uh, great expectations and, and, uh, I think the other one called data hub or something like that. But, uh, they would see that, uh, perhaps in, uh, creating a data quality squad that also does not use these things or implement them.

Interviewee

But every time there is an issue, for example, I will have to amend my code and raise an issue to them. Like, Hey guys, this is the. Query or code that that fails due to this reason. And, uh, they would follow up with fixing that upstream tracking, where does it come from, which field, and so on.

Interviewer 1

I see, I see.

Interviewer 1

Thank you. So yeah, basically you do not have, uh, time is that provided to you to implement these tools in your architecture basically. Yeah.

Interviewee

Okay. Yeah. Perfect. Yeah.

Interviewer 1

Thank you. Uh, moving on to data preparation. Have you ever measured the quality of your data and or tried to improve it?

Interviewee

Uh, to be honest, not in a, a frequent, continuous based.

Interviewee

So after facing some of these issues, I was of course filtering data that, you know, fails at any of these filters. And when I wanted to implement that, I was looking at, uh, basically the question pop up to my head is like, okay, how much data am I losing? Because they just, they don't make the cut of my filters.

Interviewee

Uh, but I didn't take it as like a, like continuously, uh, recording. How much is the, the, the total, for example, uh, uh, data that I have and how much I lose after cleaning and after pre-processing. I, I did not like put something like that, uh, in place, just one time checking on it and then, and somehow blindly assume that, uh, the ratio will be about the.

Interviewee

I see, I see.

Interviewer 1

Thank you. Um, I will jump back to data collection. I, I'm sorry, I think we have time to ask some question about it. So, um, mm-hmm. , there's three way you can collect data generally, or more or less. Uh, you can ask some people that fetch data for you. So data collectors, uh, you can use external data that, that is available, like public dataset, third party, e p, uh, or you can use maybe data that that is generated by another system, like a, I dunno, something that, uh, register the record, the trans transaction of some machine.

Interviewer 1

Uh, so yeah, there is a, these three types. I wonder is there any of these three types you have, well, three way of collecting data. Have you ever used data that was collecting, collected using one of these three types? Sorry, I'm still a bit sleepy. ,

Interviewee

uh, I, if, if I had you correctly, the last one is like data that is produced by the systems.

Interviewee

Exactly. Yeah. Uh, yeah. Uh, I, I believe strictly we have used data that are generated by our systems, uh, be it, uh, click stream data, clicks and views of users clicking here or clicking there, or generally click stream or transactions and, and bids and, uh, clicks or, or CT air records, for example. Or you can call them leads.

Interviewee

So mostly, or all of them are actually data that are gathered or collected by our systems. And we have multiple channels and, and multiple, uh, things in which these data will flow in. And, uh, as I mentioned, uh, like there are even images. There are, uh, a lot of, uh, structured data and there's also. Uh, not, not structured, but like saved in something like MongoDB, that's no synco, uh, all of, uh, all of that.

Interviewee

There is even other types that are initiatives that taken place to record certain audio, uh, and taking certain videos. We did not tap to them in terms of, um, data science or machine learning use cases, but, uh, they do exist somewhere in our systems.

Interviewer 1

Okay. I see. And so what is some of the issues you have with data that is generated by another

Interviewee

system?

Interviewee

Oh, um, I don't have, uh, a particular issue with them, uh, in terms of like, in terms of accepting it as a data, uh, the point where, Or the question that comes to my head if I'm having certain data like this, is how is it applicable to the use case that we have at hand? So, uh, for example, if we are looking into recommendation engines, we are mostly, uh, looking at what do we have in our inventory in terms of item, right?

Interviewee

So we cannot, like, involve hypothetical items or things in there. This is just the, uh, limitations of the use case itself. Uh, and when we look also on the other side, uh, of data that's used for the recommendation use case, uh, there is, uh, uh, basically there is, uh, a lot of data that we. Be it explicit or implicit data about users in the platform that makes the use of, uh, synthetic data perhaps, uh, uh, like needed or, uh, even possible at some use cases.

Interviewee

However, uh, when it comes with working with, uh, external partners, uh, when it comes to, uh, working with different, uh, programs, uh, the beginning of last year there was a program with the university and we want to give them real use cases. And, um, and everyone was thinking about, oh, how, how we can do that.

Interviewee

And, uh, I took the intuitive to basically use synthetic data. I take few, not really few, a bit more than few records and use one of the. Platforms online to generate, uh, basically as much as they want in terms of data sets for any use case and, uh, without the need. So the, the reason of it is that, for example, in this case, we have students, we want this program to be as realistic, as close to industry.

Interviewee

If we give them our own data, we cannot give it to them in the same shape. We will have to do certain things that perhaps jeopardize how this data simulate real world entities. For example, if we're talking about cars, perhaps like the prices, we will mess them up in certain way. Uh, the names of brands and models.

Interviewee

Perhaps even like hashing them or something. And that takes the essence of these use cases. And I believe if there is a collaboration with certain, uh, institutes for example, it would do, it would do the same. So here where synthetic data plays a very nice role in, uh, preserving the integrity of the data we have in our system, but giving us an alternative of data that looks exactly the same, uh, and, you know, uh, the, uh, partners or program, whatever university they have, uh, a good experience.

Interviewee

And for us, we feel safe that yeah, our exact data is not, uh, compromised.

Interviewer 1

I see. Interesting. And, and what is the platform that allows you to do that? Uh, I

Interviewee

used Gretel .

Interviewer 1

Okay. I dunno what's, but

Interviewee

I'll have a look. Gretel, uh, yeah, Gretel is G R E T E L I think ai uh, It is very nice platform. Uh, to be honest, at certain point I actually ask one of my friends that, Hey, uh, take this 50,000 rows of data from Gretel, that's synthetic data and can we figure a way how much of it look exactly like data from our system?

Interviewee

Because I look at it and I really can't tell the difference. Like is, is the price range is somewhat similar, like our distribution, is it actually totally meta price range? Like how, you know, how similar it is? Uh, it's very hard to tell on face value. Uh, but yeah, I, I use that uh, platform. Basically they have good free trial period.

Interviewer 1

Alright, that's great. Yes, I'm on the website right now. It looks i'll, I'll have look into that. It looks interesting. Yeah. Yeah. Anyway, thank you. And is there any other data quality should be missed that you can still relevant?

Interviewee

Um, uh, not really. I think there is, uh, one more of, uh, a concern in quality that, uh, perhaps, uh, so when, when we, when we think of, uh, quality N M I, uh, in ML systems or even not in ML systems, there, there's, uh, one aspect that is, uh, pretty hard.

Interviewee

I think in my understanding, and I'm not very expert, I only worked for five years that I see it a bit, uh, challenging to, to man to detect perhaps even. So, uh, you can measure things like, uh, for example, you can put validations or things that to check if certain fields that not abide, abide by certain data type rules or if certain values go out of certain rules.

Interviewee

Like for example, if you have loans or bank accounts and the age should not be less than 18, perhaps, you know, something like that that come, could come from, uh, you know, industry or domain, uh, expertise or regulations and so on. But, uh, how do we determine if something else is wrong? Because. Uh, just like logically, uh, wrong.

Interviewee

So for example, like, uh, if you have, uh, certain, let's say subscriptions, uh, for example, and, uh, and perhaps like, uh, you, you know, that, uh, let's say your, uh, your, I'm trying to use the, the metrics, uh, uh, not, uh, not g m v, it's, it's more of like basically your monthly or month over month, uh, subscription, uh, revenue.

Interviewee

Uh, it's logically to be in a certain range. If it's, if it's beyond certain range, maybe because someone just missed to make the dot and zero zero, we just kept the whole number. Something like that. Something that jeopardize the integrity of the data in a different manner. How? How? How are we, for example, able to catch something like that?

Interviewee

Uh, I find this aspect, I don't know really how to say it. I was thinking of the right word to describe that. Uh, but I can't, English is my second language. Uh, so yeah. Uh, perhaps it's like in, in a, in a natural language, uh, context, it's like the syntax is right, but the semantics is problematic. So something like that, like syntax is good, but semantically this is, this, this is not possible.

Interviewee

Uh, uh, and, and yeah, and how, how, how that can be caught. How certain things, uh, perhaps can be detected or, uh, something like that.

Interviewer 1

I see. Great. Thank you. Um, how do you evaluate the quality of. And as a reminder, quality is not only defined by the ML performance, like accuracy, iPhone score, but by also other aspect such as explainability, scalability, purpose, robustness.

Interviewer 1

Yeah. You name

Interviewee

it. Yeah. Uh, so in terms of, uh, in terms of, uh, so first, how, how perhaps can we, uh, quantify, uh, quantify this? Uh, different models have, uh, slightly different way. So for example, if we have prediction models, uh, every time, uh, they, uh, I think we have it every other day is kind of get re. It doesn't deploy, it just, we will run hyper parameter optimization over, uh, a huge number of basically, uh, varied, uh, uh, models.

Interviewee

And all of that will be recorded, but nothing gets deployed. And this way we kind of have, uh, uh, basically you can say new contenders of models with certain, uh, with certain, uh, perhaps metrics, uh, uh, com uh, yeah, metrics that we can compare with. And these metrics, uh, are not only those statistical, we also have metrics.

Interviewee

Uh, for example, perhaps it could be something along the lines of the margin of error. Uh, and we look at the margin of error in a distribution kind of way. So, uh, for example, we would be like, uh, if, if we. If we, uh, tested these models on the last two months of data, what is the percentage of data points that are falling within positive and negative 5% of error?

Interviewee

Uh, something like that. So this positive and negative, uh, uh, uh, basically, uh, five percentage, uh, of error is computed on, on a real, uh, on a real thing. Like for example, how, how much dollars does it over predict or end predict? And then we look into that distribution. So this is one. And the model that is in, in production, uh, basically we are evaluating it continuously with every prediction it makes.

Interviewee

So the moment we, we start seeing like, okay, this one started, uh, uh, basically degrade and deteriorate. We already know the contenders of models that, uh, that are better and they can be promoted into production. Uh, it this, so this kind of, uh, first, this is an online, uh, prediction or online machine learning, uh, use case now,

uh, which more of into like, uh, prediction now when it comes into other, and perhaps like this has, this is not recommendation.

Interviewee

If we have a recommendation, then of course recommendations have different approach to do it. But like, for example, certain recommendations where we use similarity based approach, for example. Now that becomes tricky because there is, there is, uh, there is no, uh, there is no easy way just to say like, oh, our, the recommendations we are generating, uh, good or bad.

Interviewee

So for those, uh, the way we, the way we do this is that it becomes an assignment for our data insight, uh, data insights team, uh, to evaluate, uh, last two weeks or last four weeks, uh, recommendations, for example, and this will be evaluated based on, uh, CTR based on, uh, even conversion. And the conversion is, if it's an auction conversion is not considered like, uh, sold or bought.

Interviewee

It's more of like bid on if, if the person bid on, because the transaction might not happen into like money transaction, uh, due so many reason. But if a person bid on something that was recommended, it's considered like, um, a hit or a conversion that happened. So this is how this happens more on a, a slightly relaxed time base.

Interviewee

And it happens over time. And from there, it, uh, we kinda accumulate insights into what improvements that can be brought, uh, and what things that, uh, perhaps in, uh, observations that we, we realize on, uh, on these things, for example, to it can contribute usually into what, uh, what the next basically iteration of this use case should be.

Interviewee

Like, what should be taken into consideration? Uh, these, uh, yeah, so like, these are two things that I have, uh, that, that I have looked into. Uh, another thing is, Uh, change of data. Uh, I'm not talking more about data drift, it's just more of like, if there has been a change into the shape of the data. So, uh, perhaps, uh, in, in, in

the car business, uh, there every car will have a brand, uh, brand name will have a model.

Interviewee

And then, um, here in Southeast Asia, we have something called variant. So the variant is, uh, every car brand usually here will have, will release few variants. So you have the cheaper one, and you have the one that is a bit more expensive, maybe because they have leather jacket, uh, sorry, leather, uh, seats.

Interviewee

And the, the wheel will have those buttons and then the higher one, and then the special edition that has perhaps a panoramic roof or something. Now, this level of granularity brings so much trouble because there are so many ways in which the people. To record these cars when as they enter the system, uh, can write them.

Interviewee

And after some time, like, you know, problems arise and the business start like, okay, let's standardize this, uh, based on a certain data set, for example. Now this new data set might not match the old dataset that we have. There are certain changes, but all the models that we have has been trained based on the older data.

Interviewee

And now how do we make this switch? Uh, it, it's a whole, uh, thing in its own, uh, people will think of it in many different ways. Uh, uh, to be honest, and especially for our prediction use cases, that becomes key because the first thing is like, so you're just telling me that I'm gonna lose two, three years worth of data?

Interviewee

And I have this new one, can we match, like, you know, say, Hey, this is in the old one, written this way, it's equivalent into this in the new data set, or something like that. So that, uh, contribute into, you know, disruption. There is even, uh, there is no easy way even on the use cases. I, I have already hand over my project spec, basically, but I put these questions that we don't understand the impact of this change into our recommendation engines.

Interviewee

We have been, for example, processing the features in a certain way, uh, encoding them in a certain way now because things have changed. What does that mean? Uh, are, are we experiencing certain, uh, basically even drift in, in, in, uh, how, how good or. How relevant our predictions are, are were questions of that thought.

Interviewee

So, yeah, I don't know. Maybe this is just, uh, uh, an each use case. It doesn't happen always, but might happen as well, but it brings a real concern. Uh, I have a, a discussion, uh, also with, um, authentic person. Uh, and I, I asked him, he, he's from the Philippines. They have authentic there. Uh, and, uh, I mean like they have, uh, authentic startup there.

Interviewee

And my question was like, okay, so you have been basically having a credit scoring for the last two years, then the pandemic happened, and so much change into how people work, how much they make, how much, uh, you know, the cost of, of living as well. And now everyone is start kind of going back to work. How do you, I mean, like, You know, which data do you trust?

Interviewee

How, how do youve verify and validate the impact of this thing in your models and your assumptions? That took perhaps like a year and a half until you reach a stage where you are happy. And he explained to me that they have certain, they have certain basically, uh, financial, uh, tools, uh, perhaps something called P C R or something like they have certain things that he's saying.

Interviewee

If my models will drift beyond this value in this financial metric, I will not trust it. I will fall back to something else. Uh, in industry I'm in, uh, there is no such thing. It's not as regulated as FinTech. So we kind of like, have perhaps research questions and depends on the individual and the prioritization.

Interviewee

If, if people put effort and, and, you know, look at these, uh, uh, questions. Sorry, I'm taking quite a lot of time to just

Interviewer 1

No, no worries. You mentioned a lot of interesting point, uh, so I'd like to go back on two things. You mentioned the latest, well, the first one is your, your friend who works in a FinTech.

Interviewer 1

Uh, basically a prediction of a model will not be trusted if some metric is, um, abnormal or below, above a threshold. Right? Okay. Mm-hmm. . And, um, and how, how is it presented to the user? Is it, uh, the model cannot be trust at this moment or they fall back to some other version to predict, uh, a value?

Interviewee

Yeah, so, uh, so the reason I asked him, because I was contributing to a FinTech startup here that is working basically in stealth mode, building a credit scoring.

Interviewee

So I asked him from the point of view of, uh, how do you trust it? A credit score model is not something that's always, or at least here from what I see, it's, it's more operational. It's more something that tells the business, can you trust this loan applicant? Are they safe or not? So it's not always kind of disclosed, uh, uh, to, to the user.

Interviewee

So, so the impact, it's more like, okay, so there has been this massive change in the data, in people, in lives and everything, and you have a lot of assumptions in your model, but how can you, how can you test, how can you ma tell the business everything is still good and intact and you can carry on, uh, trusting, uh, basically the scores that this, uh, model is producing.

Interviewee

So, so that's basically what he, uh, brought forward is that there is a, a p, uh, something called P C I. I might be wrong by the way. I just heard it, uh, in some Starbucks . So he called, he told me like there is, uh, basically the concept is benchmarking. He told me that you have to benchmark. And he explained to me that

how he's benchmarking is he is measuring the predictions that he has, uh, in corresponding to this thing called pci.

Interviewee

And based on that, he is like, this is, uh, this is kind of, uh, his compass is like, should I go with this model? Should I stop and fall back to something else? And, uh, something else is, could be like, uh, you know, a credit officer, like literally just fall back into a human being that makes the decisions or basically anything that is more trustworthy but less automated.

Interviewee

It it, that's like the kind of the way so, I, I even googled this a lot, uh, about benchmarking, uh, is it called benchmarking actually or something else? Uh, yeah. Uh, I don't, I know, I know what's having it. Yeah. Okay, then, uh, okay, good. Because I started thinking like, did I hear correctly? So, yeah. Uh, basically like that's like the, the way, and that's the practice, uh, that, that people would use more.

Interviewee

And, uh, yeah, it left me with question like, okay, how does that apply to the use cases that I have? How do I implement that? But yeah, this is, uh, basically, uh, what, what I can say about this. Yeah. Okay. Great. Great.

Interviewer 1

And, and the pci, uh, yeah, the pci, is it another model? Do you know if it's another model or, or, or if it's, uh, another process to generate, uh, predictions.

Interviewee

So from what I understood, the PCI is something like, uh, something regulated. It's not even created by them. It's like, uh, perhaps like, uh, it's like an index I think from a bank or something. And, and, and they will look, uh, into it as like, this is the base mar uh, this is the baseline. Uh, or something like that.

Interviewee

Yeah. Sorry, I'm, I'm very shallow in this area. Uh, I did no work on it. I was just very curious to talk with this about him because I was contributing to a startup that was also like trying to build a credit scoring.

Interviewer 1

No, it's pretty interesting. And, uh, the last thing, last thing I will ask you regarding what you just talked about is, um, you mentioned that sometimes the data collection process changes and it makes your data worthless.

Interviewer 1

So for example, you mentioned car with different variation and if someone change away, they, uh, enter the data. , it makes every, every data dataset, well, every data in the past use less. Um, how do you address this problem? Do you just throw out the dataset or do you have some tool that say, oh, maybe this correspond to this, um, or you talk with the people and they tell you how to transform the data?

Interviewer 1

Yeah,

Interviewee

yeah. So, uh, first, like we, we can't, we can't afford at certain use cases, we can't afford to throw the data that, uh, that's like, that's a complete nightmare because, uh, it's basically a not option. Uh, and so we have to find a way. So the way we found is that, number one, uh, trying to do matching and train a model and try to see like, okay, how bad are we

Interviewee

And from there trying to explore like, Uh, a form of like manual labeling perhaps. Basically trying to match and saying, oh, this is equivalent to this. Something like that. And try to see like, uh, okay, using this, uh, how much weeks do we need, how much data we can generate? And, and, and, uh, perhaps spin an iteration basically to do that and look at how does that impact the model?

Interviewee

And, and then that kind of sets a tone of like, okay, are, are basically, are we very, very, that basically destroying every, uh, every SLA that we have, or are we very close and, uh, it's a matter of adoption basically on our. And all of this while af after something like this is, uh, uh, you can say officiated and all the businesses

using it, uh, we are in the trenches, like still like, oh, wait, wait, wait.

Interviewee

We'll use the old model still. And we are training this new model and we are trying to see how close we are. So it's more of, uh, yeah, trying to work, uh, on that. Uh, and uh, uh, to be honest, sometimes like you don't, you don't have so much saying like, no guys, stop this. We are 50% bad. It might as well consult your cat in this predictions and it'll have 50% correct.

Interviewee

Uh, like you, you don't have like, uh, that, that power. So it's more of, uh, figuring out a way, figuring out a way to like on, uh, like I said, like the first thing is just like try to match and on this small data that perhaps, uh, O only last for like two months compared like two 20 months that we have, can we generate something decent and how far we are?

Interviewee

And then we look into ways that does an increase in the data co, uh, correlate with an increase in the predictability of the model. And we try to, at this level, we try just to get closer to our, uh, standard, uh, in other use cases we like in the recommendation, for example, it did not matter, uh, much because the recommendations first the way the recommendations are evaluated is different and these recommendations does not, does not, um, Does not directly, uh, correlate into lose of money or anything.

Interviewee

They're mostly like, uh, uh, responsible for, uh, uh, carousel. Basically how the carousel appears, it's more of personalization. So here, like the severity is less, and, uh, because on every, like for example on the C2 B, we have auctions and every auction has a specific inventory. So like, we have to train a model based on this inventory.

Interviewee

So every time we train, uh, a model, every day, every session. So if the data changed and we can read a consistent data for that specific session, there is no problem. In the operation side, in the quality, what are the ramification of this change? That's like a whole, uh, different side of it. But yeah, this is how, uh, how we went

through, uh, this use case.

Interviewee

Uh, we were actually impressed that, uh, we were able to generate, uh, basically a similar level of performance in a lesser data. Uh, but we were like, we wanted to be sure. So we went on, uh, uh, actually we have to use the whole team, everyone to, to work on this kind of manual labeling. Uh, give everyone just like 20 lines and then you can end up with yeah, uh, 2000 new rows or new data points out of nowhere.

Interviewee

But, uh, generally, like as much as it was a concern, uh, somehow, like, it, it, it, uh, it worked. I think it worked also because, For this use case, seasonality is different. Uh, if, I don't know how is it, uh, in your side of the world, but in this side of the world, if you want a car today, especially certain models of cars, uh, you have a long, long queue, like months, maybe six months, you will not see that car yet.

Interviewee

And what this, what does this do is that the secondhand car market is booming and we are in that market. So like there isn't so much fluctuation and somehow like our models were able to perform on a similar level because they did not have, uh, there, there weren't basically high seasonality in, in, in, in a slightly longer timeframe.

Interviewee

And that kind of lessened the impact. Perhaps it's all just theory. It's just how I imagine it.

Interviewer 1

I see. Thank you. I'll have two more questions for you, uh, because we, we are already, uh, I don't want to spend too much of your time. Um, so what are the challenge you have encountered when deploying or maintaining machine learning software system?

Interviewee

Uh, I think the first, uh, the first issue that I have is, uh, uh, basically understanding, uh, understanding how, how things will go in terms of load, in terms of, uh, what's, uh, what basically scalability should I look at, uh, how, what should I, uh, think of it and, uh, uh, to especially like for, for me, I was involved like in, in the whole, uh, thing.

Interviewee

I deploy my models myself. I have a limit on, on, uh, uh, what technologies that I know out there. Uh, I might know certain technologies, uh, theoretically, But definitely don't know how to set them up, uh, or something like that. Uh, so, uh, that always, uh, come, uh, as a challenge and, uh, perhaps until now I'm lucky that things somehow work.

Interviewee

Uh, fine. Uh, so along the, the mostly it comes from architecture and software engineering challenge. Uh, in data science. We are not like software engineers. My software engineering experience is, uh, being an intern for few months. So, uh, not so much. Uh, Uh, on my first use case, one of the problems that I have, uh, which almost like made the world close of me, is that, uh, they were complaining.

Interviewee

My p i is taking very long to respond. They're like, this, this response time cannot be in the homepage. I'm like, if you put my, uh, carousel in, in the last page, and then like, how do I prove there is even traffic or adoption or anything? Nobody will look at it. And, uh, yeah, so like I go to talk to software engineers and then I analyze that, oh, there is, uh, of course something like multi-threading or something, but now I'm talking about endpoints, like how to make it fast and there are more concepts and more things and, and it's like, oh my God, just like, I, I am not even sure I can make this work.

Interviewee

Uh, so that, that's basically one of the challenges that I have. Uh, funny just to mention, I was using flask. That is the time where I discovered something else called Fast, a p i. And to be honest, fast, p I just excused my lack of knowledge in a lot of software engineering concept. It was so fast and I didn't have to learn so much about async and, and all of these things.

Interviewee

It, it helped me doing them in a way. It's like it empowered me despite of the lack of, of, uh, of knowledge on, on, uh, deepest software engineering concepts. Uh, and that's how it went by and, and from there until today, I stick with flash. P i is the tool that's empowered me. , uh, uh, another challenge that comes, uh, or along the side, like when you, uh, because I mentioned like there is software engineer and architecture.

Interviewee

So when it comes to architecture, for example, uh, there are things along the line. So for example, I'm dockerizing my app, like, is this enough? Can I just put this dock and I'm running it in certain server and, and, and that's it. For example, I went on doing that, uh, more often, but later on, for example, I get to, uh, questions be like, uh, how much traffic do we have?

Interviewee

Uh, is it like, what if you have basically a huge traffic at a certain point? How would you, uh, do some scalability? And like I mentioned, like, uh, somehow I read more about, uh, ES, but like, okay, ES can do X, Y, Z can solve somehow all my problems, but I can't myself, uh, basically, uh, you know, take my things and put them into Kubernetes, for example.

Interviewee

So that's the architecture kind of challenge. Uh, the other side of, uh, the, the challenge comes from, uh, basically sustainability and, and, uh, Yeah, I just got three points. Okay. So I talked first just about like software engineer and architecture. I think these two points are two sides of one point. Uh, the second point is about, uh, oh, I forgot that.

Interviewee

Okay, let me take the third point and get back to that. So the third point is more into, uh, uh, basically how do I, uh, how do I prove impact to the business? For example, uh, I, there are times where I feel like I was lucky people understood , uh, and, uh, perhaps not understood me, but like maybe they did the homework as well and they can agree with me at certain things.

Interviewee

Uh, especially something like, uh, uh, so, uh, especially on, on the recommendation engines, because I built more of these use cases. One of the challenges like proving. That. Uh, so one thing is that people perhaps generalize this thing, do your recommendation work and to means like, what, what do you mean? It work?

Interviewee

It generate recommendations every single day. If you, if that's the work for you is like, it is working, but when it comes like to business, does it work for business? Does it generate impact? Now here comes like the most, uh, the most painful part for me at least. Uh, in one episodes of the experience I have where, uh, I, I follow up with people from the industry, I kind of follow up with like, okay, how do they measure, uh, impact from, from, for example, recommendations, but.

Interviewee

It's in a way is like, uh, uh, perhaps the leadership is not with me on the same page, or my manager doesn't feel like fully confident and it's new for him. Maybe he is not exposed to, uh, uh, to that side basically of the business so he doesn't feel fully confident that, okay, this, we can take this as metrics.

Interviewee

That to me is the, uh, most devastating challenge because, uh, it's very, it's very hard. You are, you are left basically with, uh, a lot of business metrics from different sides. You are trying to say adoption wise, uh, uh, adoption wise, uh, top funnel wise, whatever. Basically that is on the top funnel o o of the business that you want to prove.

Interviewee

And then you kind of try to materialize these things into dollars and say like, oh, okay, so because we have this much of C T R. Uh, it's increased, for example, the top funnel by this much. So we are responsible for this much, for example, of, of the revenue or something like that. But, uh, does not necessarily transcend into, uh, the, the, the rest of the leadership or, uh, or beyond basically the top line manager that, uh, that, that is a very, uh, challenging, uh, a very challenging thing, uh, if basically proving the impact of your model.

Interviewee

Uh, differing cases are, are different. Uh, but, uh, and, and basically different, uh, people perhaps processing it differently, uh, to mention in, in the semantic search, uh, use case, for example, the way I try to prove that my model is better. So my, uh, semantic search engine does not really do the whole search.

Interviewee

It just, it takes the turn. Uh, sorry, the hundred search results and it re ranked them says, what should be the first, what should be the second? And the comparison that I did with the benchmark is that I will look at the GMV of the top 10, and I will look at the GMV of the top 10 on the other side. And if these top 10 results has, uh, a better price, for example, or have higher, uh, ing basically from clients, then we say like, okay, it seems we are hitting the, the, the right button in here.

Interviewee

Uh, but on, uh, under recommendation, it'll go in a different way. Uh, and even in, in the predictions for example, uh, how it's always that challenge that how, how do you translate this to business? How can you go beyond your. Statistical metrics into something the business feel or speak. And, uh, that is challenge.

Interviewee

I don't know if that, if why is part of the question ? Uh, but is it, is is the why, uh, are you interested to know why is that challenging?

Interviewer 1

Why it's challenging? Yeah.

Interviewee

Yeah. Go for it. Yeah. So I, I, I think in my humble opinion, uh, that is challenging. That is challenging because number one is, for example, I, I, I come from, I did some engineering for two years and then moved to computer science.

Interviewee

I don't have business background. And I think even if I was a business student, I don't necessarily have the business language of, you know, looking at things in terms of revenue and a R R and G M V and so on. So, Uh, it, it is very challenging if you don't have that, uh, kind of, uh, mentorship and guidance that, oh, this is how you actually should do it.

Interviewee

So, to me, I think this is, this is why, and this is number one, is like, uh, it's that, but even after you are exposed to this, it doesn't mean it's not gonna be challenging because, uh, for me, I thought, like, I somehow figured out how to prove to business the, the ROI or the impact of my, uh, machine learning models.

Interviewee

The challenge was like, uh, at a, at an episode, my direct manager does not share. Does not share the same, uh, uh, view or understanding as me. So, uh, he, he, it's not that he wants different metrics, it's more of he is skeptical if these are the right metrics and he doesn't have the business acumen that he'll be like, I want X and y, uh, metrics, for example.

Interviewee

So it end up in a way you kinda think that as a machine learning engineer, you kind of know because you are familiar with the use case and you are trying to prove, but your proof is not, uh, considered or, or taken forward. Uh, these, these two challenges I think are, are the hardest and, uh, and, and, uh, uh, basically why, uh, uh, or these are two reasons are why, uh, the last thing that is challenging.

Interviewee

I just got the idea. , uh, it, it is basically in, uh, sustainability. So when it comes to sustainability, uh, handing over to other people becomes very challenging. Uh, and, uh, in my experience, uh, that was actually easier with certain teams, but harder on certain teams. And it comes to the, the, I think the background people have and their real, or, or you can say their skills and background.

Interviewee

So if people are familiar with, let's say or kind of problems, and that is the data science they are, uh, familiar with, and then they move into something else that is more of predictions based and clustering and so on. Uh, Passing, uh, such, uh, torch is very hard. Uh, they, they have, uh, basically lack of context at certain things.

Interviewee

Lack of familiarity of, of how these things work in production. Uh, uh, it becomes very challenging. You have to document everything you have even to document. Uh, often you see, uh, F a Q or frequently asked questions. Uh, I have something similar, which is like frequently, uh, uh, frequently occurring problems.

Interviewee

I always add that to my documentation because I just want people to know like, Hey, look. These three or four, they are always your friends. Every time if something break up and it's one of these, here is the recipe of how to solve them. Uh, but if it's not, then uh, you know, it's, uh, it's uh, kind of harming, uh, the, the sustainability of, uh, of the challenge.

Interviewee

But yeah, sustainability is the, definitely the, the third challenge. And, uh, yeah, uh, three points. I see, I guess. Thank you.

Interviewer 1

Remain interesting. Alright, uh, so I'll ask my last question. In your opinion, what is the most pressing quality issue researchers should try to solve?

Interviewee

Uh, oh, I think the last. Uh, the last issue that I mentioned to you where the syntax of the data is correct, but the semantic is not. I think that's very important. Uh, I have ventured in my own and I have a client and the product manager have always asked me, uh, how do we validate that the metrics that you built for us are correct?

Interviewee

And I always stuck there because I'm like, it's a very good question, but like, uh, I really have no answer.

There is no such tool, uh, or to my opinion, like there is no such tool because he's, he is referring more to the semantic of, of the, of these KPIs. So I think that's one, that one, uh, challenge there, that that is very.

Interviewee

I, I'm not sure, maybe like, I'm really not aware about tools out there, but I think this is one, uh, definitely, uh, one of the challenges that are there. Uh, when it comes to, when it comes, uh, I think when it comes to practice as well, uh, usually in, uh, especially like for problems like imaging and so on, uh, I think the, the, the practice there are somewhat like kind of, kind of lacking or perhaps in my experience, for example, how, uh, how uh, uh, perhaps interpreting or reasoning why certain, uh, misclassification or mis annotation in object detection problems happen is very, is very hard.

Interviewee

Uh, and we have to look at the image, like is are there ways in which, uh, You know, uh, I don't want to say explainable ai, but perhaps explainable AI for image, I think is, is a, is an area there that can, because it can reduce the, the time that, uh, you know, usually people, uh, spend to, to figure out exactly why a certain prediction or certain, uh, error is arising.

Interviewee

Yeah. Uh, I think, uh, uh, these, these are on my experience. Uh, I hope like, uh, uh, I, I don't know. Are you publishing the, the results or something, uh, in a way, like what are the most pressing problems? Because I don't know actually if companies across the world, like facing a similar issues or certain issues can relate to only certain parts of the world.

Interviewee

Based of perhaps data practices and regulations there. So yeah, I'm, I'm, uh, that, that's my reason. And, uh, yeah, I, I don't have . Yeah. Yeah. No,

Interviewer 1

thank you. Everything you said was really interesting and I, I'm sure it'll be useful for 30. So, uh, really no need to worry, in my opinion. It was great. Yeah. All right.

Interviewer 1

Uh, you mentioned a little something, uh, I'll, I will quickly go over it like one minute. Uh, so you, the first issue you think that researcher researchers to try to solve, uh, you mentioned, you mentioned detecting, um, VA like values that are abnormal based on the semantic, right. And you also mentioned, uh, metrics that do not mean anything to, uh, a customer or your boss that is in business.

Interviewer 1

Am I correct? Uh, yeah,

Interviewee

perfectly. Yeah. Okay. Perfect.

Interviewer 1

Thank you. All right. Uh, so that's it for us. Thank you. We went way over time and uh, I think it was really interesting. So thank you for your time.

Interviewee

Thank you very much. I wish you all the best. Same for you. Thank

Interviewer 1

you for your

Interviewee

time. Yeah, it was very comprehensive.

Interviewee

Yeah. Thank you. A good luck. All right. Have a good day. Bye. Yeah, thanks. Bye. Have a good day.

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