

Interview 15 - Amin

Interviewer 1

Okay, perfect. Thanks.

Interviewer 1

All right. So my name is, as I told you, my name is Pi. I am a master's student, uh, Rashad, that you can present yourself if you want. Um, I'm

Interviewee

also a master student at and the same lab.

Interviewer 1

Cool. Alright. And IIN is, um, I serve research a. Cool. All right. Uh, so, um, if you want, you can present yourself, you can give us some background information, your current position. Sure. Experience.

Interviewee

Yeah. So I'm X. I, uh, uh, I was a data scientist before I, my current position, so I'm currently a statistician at Company X.

Interviewee

Um, I. Here as a statistician for the past three years almost. And before that I was a data scientist, um, for maybe four years in the private sector. Um, so I've put machine learning products into production, both in the private sector and then now in the international organization. Yeah, so my, my background is in economics and econometrics.

Interviewee

Um, but I've done a lot of machine learning, machine learning work since then. Yeah.

Interviewer 1

All right, great. Uh, so we we're glad to have you today. Thank you. Uh, so I will just describe quickly what are the goal of the interview? Mm-hmm. . So we want to develop a, a catalog of quality issues, as you know, of quality issues and machine learning software.

Interviewer 1

Uh, so the first thing is what is a quality issue? You have two words. Quality and issue. Issue usually is pretty, pretty clear. Uh, but quality, uh, sometimes it's a bit more vague. So I will give you an example. Uh, let's say you have two system that accomplish the same thing, but you're able to say that one is better than another.

Interviewer 1

Well, what you usually referring to is quality, right? And machine learning software system is any software system. Uh, that has a machine learning component in it. So really simple. All right. Uh, so just before you begin, do you have any question on, on

Interviewee

anything? No. I guess I will become more clear as the questions come.

Interviewer 1

Okay, perfect. And don't hesitate if you have some question. Yeah, ask me. There's no words. All right, so we'll start with a general question. Oh, and by the way, there, there is like 20 or 25 question, uh, just to give you some. Uh, so the first really open ended question is what are the main quality issues you have encountered with your data model or software system so far?

Interviewee

Okay. A big question. So there's a lot of components in there. So the quality issue with the model, with the data, Are two different things. I can speak quickly. So the, I guess the context that I'll most pull from is our, our Now casting, um, product that we have at anta. So what that is is, uh, uh, forecasting global trade and gdp.

Interviewee

And this is done using, uh, neural networks. So that's machine learning component. But of course, as you guys know, by now, there's a whole. Component in terms of the data collection, what data goes into the model, as well as the final presentation. How do you present the results of the model? How do people interact with it?

Interviewee

So briefly on the, um, data collection, quality issues, side of things, um, there's a lot of, let's say, yeah, change that can happen. So what we've found is that a lot of the data sources we would use in. Beginning of the project either change their, um, let's say how they're hosted, so it go, they change their api, they change their, um, the format of the files that they post, that sort of thing.

Data-unstable-provider

MLP-unstable-external-dependency

Interviewee

So that's at a very basic level, uh, quality issue where you don't have necessarily consistent, um, data presentation and avail. And then that can go into other areas as well. Some series have actually been either rebased or they've been redefined. This is especially true in the economic context where, um, most of our data come from where the, uh, statistical office will redefine that, um, whatever I say, the consumer price index and have a different basket.

Data-unstable-provider

MLP-unstable-external-dependency

Interviewee

And sometimes it may not be comparable. So there's the element of just the data availability, but then of the actual. Content of the data itself being, um, comparable over time or not. So these are a couple of the problems that we've faced and have to go back into and adjust. Either, uh, retrain the model, see if that that variable still works.

Data-unstable-provider

MLP-unstable-external-dependency

Interviewee

Um, drop the variable, not drop the variable if the new, new delivery mechanisms. Complicated or inconsistent. We found that, you know, they change it every two months and it's too often for us to go in and reincorporate it every time. So there's um, that element where the data itself may be good in the model, but

there's that extra consideration of, you know, the data don't just arrive to you magically in an Excel sheet.

Data-unstable-provider

MLP-unstable-external-dependency

Interviewee

You have to go get it. And I definitely have dropped variables from the model just because I no longer either had confidence. Didn't want to deal with the delivery mechanism that that particular institution used to deliver their data. So that's on the data collection side. Um, there's also elements of, we do a lot of transformation to the data as well, seasonally adjusting it, um, normalizing all that kind of thing.

Data-unstable-provider

Data-unstable-provider-fix

MLP-unstable-external-dependency

Interviewee

And sometimes the, especially during Covid, we experienced this where a lot of the, um, some of the changes we also work on. Growth rates, so not the actual values themselves. That during covid some things went literally to zero. For instance, we had Spanish tourist arrivals as one of our input variables for services trade.

Data-containing-outliers

Data-noisy-features

tabular-issue

Interviewee

And then of course, the next month you get infinity because to growth in zero, it's infinity. So there's that element of, uh, data quality as well in during turbulent times. Um, you have a lot of missing. Is another one where they will just skip periods, observations, they will have a long delay, uh, in publishing the latest data, all that sorts of things.

Data-containing-outliers

Data-noisy-features

tabular-issue

Interviewee

So a lot of, uh, of what the now casting, which just to be clear now, cast is kind of, it is just forecasting for the current period. Um, so say that we wanna now cast GDP for the fourth quarter. That's what we're doing right now, even though the fourth quarter is ostensibly happen. You still forecast it because the, um, first of all, it's not finished yet, and second of all, the final figures won't be available for like another three to six months, depending on the country.

Data-containing-outliers

Data-noisy-features

Interviewee

So you want to know that figure before, you know, six months from now when people actually publish it. That's why it's now casts not forecast just as some context. So a lot. In this domain deals with dealing with the missing data, dealing with the, the missing data at the end of the series, as well as intro series within the series.

Interviewee

Um, for instance, some Chinese statistics don't publish the month of February because it's such an anomalous month because they have the Chinese New Year, and it's just a lot of, um, weird behavior. So a lot of things have to be robust to these types of issues. Then in terms of the modeling itself, that's a whole new, um, task where you collect your data, but then once you have it, how do you, you know, generate your predictions.

MM-drift

Interviewee

Um, so that obviously goes through a model selection process, variable selection, hyper primary tuning. And those are not necessarily, there's no quality issue. I'd say in that element of it. You're working with your complete data set. But what can happen is, um, over time you can get model models getting stale.

MM-drift

Interviewee

So relationships that they were picking up on are modeling become weaker over time or ceased to exist. We're kind of doing this now in some regards as the, you know, we have potential structural breaks with the, the Russian economy to no longer being as intertwined with the German economy. So historically what may have worked really well for, uh, predicting Russian outlook based on, you know, German manufacturing et cetera, is no longer necessarily the case because they have this, um, a new different structural relationship that the historical data will not reflect because it was different in the past.

MM-drift

Interviewee

So you have these elements of model drift as well. So there's kind of two ways. Address that one is by, um, continually retraining your models. So, uh, in the beginning of our process, we were modeling the data, creating the model, training it at one point in time, and then just using that for period quarters and quarters and years afterwards without retraining it.

MM-drift

MM-drift-fix

Interviewee

Um, but then we've adopted now a kind of a rolling. Training, uh, approach where every quarter you'll train a new model for that particular quarter. So that's one way to, to get ahead of models becoming a bit stale, where it doesn't solve everything because you can have new variables that become relevant and ones that become completely irrelevant, but it at least will, uh, capture the latest relationships.

MM-drift-fix

Interviewee

In the data if something's kinda changed and the intensity of their relationship, et cetera. So just having all of the latest data model trained on the latest data is one element and the other element. So that's kind of a, that will, let's say, prolong the life of your model, but it'll not be forever because we know things change.

MM-drift-fix

Interviewee

So the other element is going through a a re, basically going through the model selection process again on a regular, semi-regular. Just to see if there are new, uh, developments, new variables, new, just a new model that could potentially do the job better. Things have changed, and that's kind of like a auditing if you'll, of the, of the models in terms of their specifications, hyper parameters, the variables that go into them on a semi basis, just to make sure that you still have, um, the best performing model to your knowledge.

MM-drift-fix

Interviewee

Briefly, those are some of the issues, quality issues that I faced in, in generating, um, this product.

MM-drift-fix

Interviewer 1

Yeah. Thanks. There's a lot of information to, that's great. Thank you. Uh, so you mentioned some issue with data collection, right? Mm-hmm. , um, Where did you get your data from? Was it like a, did you do web, web scripting or was it like an API that you paid for and you received data in

Interviewee

exchange?

Interviewee

So the input data is almost, I think actually all publicly available, um, from. National statistical offices from international organization like the O E C D or the UN, or imf, et cetera. These all have APIs. So we're using APIs for all of those. Um, some things get more specific, like Hong Kong Port Authority, um, that's not quite web scripting, and they will post a, um, an Excel file with their port calls every.

Interviewee

It's kind of just paying that file, taking the latest information. We did use web scraping earlier on, had two issues with that, which are maybe a little bit specific to our use case. But, um, so for our use case, we need a long history of the data, um, because a lot of the things that we're forecasting only occur or are recorded every quarter or every year.

Data-collection-external

Data-collection-web-scrape

Data-unstable-provider

Interviewee

So you don't have too many observations, so you need, uh, pretty long history. So with the web scraping, the first problem is that you often don't have a history or a time series of the data. What's fine to go grab a snapshot and dump it and put that into a model, but we really need to know that, how it stood in the past.

Data-collection-external

Data-collection-web-scrape

Data-unstable-provider

Interviewee

So, um, if, if the source itself doesn't already provide historical information like that Hong Kong authority that she has the history. So it's not a. But if we were to try to hit, you know, the port authorities, um, website every day because they, they write it out, let's say, well, how many ships came? That wouldn't work because we would only have the data starting from when we began collecting.

Data-collection-external

Data-collection-web-scrape

Data-unstable-provider

Interviewee

And that's not really gonna cut it. You have matter of weeks or days over time, it can become a longer time series, but you have to be really consistent with it. And, um, in our experience, websites changed the two. Much too frequently for that, for you to have a lot of those running at all times. And secondly, it's um, a big job to schedule all these tasks and have them consistently and maintain them all, basically have them consistently generating and outputting data.

Data-collection-external

Data-collection-web-scrape

Data-unstable-provider

Interviewee

Um, that became too much of an overhead. So we really moved towards really focusing on APIs from organizations who produce data as a product, let's say not. Scraping data from, from websites. Uh, I see. Yeah. That's Do we use scraping for? Yeah. And yeah, that's, that's about the relationship with scraping in my experience.

Data-collection-external

Data-collection-web-scrape

Data-unstable-provider-fix

Interviewer 1

Yeah. Correct. Right. And when you use an api, I guess sometime you have issues how, I mean, you cannot really prevent this issue, but you can detect that there is problem with the data you fetch. Yeah. The automatic system to detect it, or usually it's done manual.

Data-collection-external

Data-collection-web-scrape

Data-unstable-provider-fix

Interviewee

So we have, uh, uh, because, so we collect hundreds of series, so it's not really feasible to manually keep an eye on all of them every week.

Interviewee

So basically, uh, every, so this, this process to update the data and update the models happen once a week. And as a part of that process, there's an error logging system, um, where you can scroll through and see the status of each of those calls in each of those. Processes and then you can just, you know, filter what, what failed this week.

Data-unstable-provider-fix

Interviewee

Sometimes it's just anomalous and the, um, that server was down at that time and it'll come back next week.

So you can compare the history. Has this failed three weeks in a row or did it fail once and then it worked again next week. So that's the mechanism we have in place to monitor that status. And if something does break or it hasn't produced in, in three weeks, then we can go in at Deb.

Data-unstable-provider-fix

Interviewee

Okay. What's breaking? Ah, they changed the format of that file or, oh, they changed the name of the variable, whatever, and fix it. And then, um, have that, that data source produce again. Okay.

Data-unstable-provider-fix

Interviewer 1

And the monitor monitoring, monitoring tool use, is it in-house or it's something open source?

Data-unstable-provider-fix

Interviewee

No, it's in-house. It's a, it's an element of the code.

Data-unstable-provider-fix

Interviewee

Produces the, the data. It's just another function in there that, and write the status of the, of the call to this blog file, essentially. So it's just part of the process. Okay, great. Thanks. Um,

Data-unstable-provider-fix

Interviewer 1

have you ever measured the quality of your data and or try to improve it? .

Data-unstable-provider-fix

Interviewee

So what do you mean by the, okay. So I think the quality of the data is a bit less relevant for our use case because, um, it's already vetted, it's published by NSOs, national Statistical Office.

Interviewee

It's published by organizations, so that is really done by them. So we don't have to, I don't have to go and verify Brazil's statistical output. Definitely is more relevant for the web scraping we did early on, which we dropped. Um, because you may not have throw an error, but you may start picking up the, um, letters of the after the figure, which kind of happened.

Interviewee

And so it's, it says, oh, it's all good. But then the data is actually like, kind of nonsense, nonsense. Um, so, so for us it's not, now that we've made that switch, it's not as much of an issue. So that's how we. We do it.

Interviewer 1

Okay, great. Thanks. And is there any other data quality issue we missed that you'd consider

Interviewee

relevant?

Interviewee

I don't think so. I mean, we, I mentioned missing data and I think that's, uh, an important element for our work. Um, but yeah, in general with that machine learning need something that can handle it. You're never gonna have complete perfect data. So, you know, either fill it with the means, drop the row, whatever.

Interviewee

But it's, uh, it's very relevant. All right,

Interviewer 1

great. Uh, moving on to model evaluation. How do you evaluate the quality of your model? And, uh, as a reminder, quality is not only defined by ML performance, but there is other, other aspect such as, uh, scalability, explainability, robustness. So,

Interviewee

yeah. Yeah, for sure. So I'll start with the easy one, which is the, um, say the um, assessment parameter.

Interviewee

So we usually use root means square error for the predictive accuracy of the models. We use it on an add sample test set. Um, since we're dealing with time series, it's a bit more difficult. Normally you would do cross validation, um, Fold, K fold Cross foundation we're, uh, assessing so that you have more robust results.

Interviewee

But in our case for time series, it's not usually possible cause the, um, observations are not independent. They have a time component, so you can't assess the past based on the future. So, uh, we have a rolling, uh, a rolling forward window, so where you just expand. Test period rolling forward. Uh, so you had a couple folds in that way.

Interviewee

Then you can, um, assess the model performance. Uh, so we also will do different test periods like, um, you know, 20 years apart, et cetera. So you can see, okay, this model work in the eighties, did it also work in the two thousands or has there been a bigger structural change where we, um, where that doesn't work?

Interviewee

Um, and always the kind of guiding principle is the, the rooting square air or mean absolute air. Um, depending on the nature of the series and how big its, um, fluctuations are in terms of the, uh, scalability and actual com computational performance, that's for sure an element. Uh, we need, we don't have a lot of, uh, resources for cloud computing, for servers, et cetera.

Interviewee

So, um, the methodologies need to work on a realistic time scale on kind of home hardware. Uh, just think a mid-tier tower or laptop. Um, and that was a consideration. It's not too much of an issue because when you're working with time series, you often don't. Unless it's high frequency, like minute or second observations, you don't have a ton of observations under, you know, not more than a couple hundred.

Interviewee

So it's not frequently a big problem. But if you get into very wide data sets, it can slow things down. And some of the methodologies that we use, um, that we tried, let's say, tested out before, that was a factor of consideration for these take, you know, 10 times longer to estimate. Still realistic. But that is an element in especially the model selection phase, where you're gonna have to be running many, many models where, you know, it's a difference if we do, uh, a model reevaluation that takes a day to run versus 20 days to run.

Interviewee

That can be the difference between something being feasible or not. Um, so that's consideration and the methodology used now it's quite performant and it's, um, relatively fast. It's not issue, but we, we have considerations for it when we're kind of now expanding it to potentially many, many more hundreds of, of potential target series where those scale issues, um, become more prominent.

Interviewee

When you're doing model inspection four, one model, the same approach may be feasible, but when you do it 300, it no longer is feasible. And that's, uh, an element of consideration. Um, robustness is, is, it goes as what I was talking about before where you kinda just periodically reassess the models and rerun the exercise to make sure that it's still performing well, that, uh, another potential specification hasn't, you've surfed it.

Interviewee

Uh, that's kind of the, the way to go about that. Covid also provided a great opportunity for that, where, uh, a. Classical methodologies performed perfectly fine when everything was normal, but then huge anomalies 2020 to give kind of gibberish output. And that's, that's a really nice, easy robustness check.

Interviewee

Um, right away you say, okay, if we can't handle covid, then um, we won't use it because we know that it's not gonna be equipped for the next big hope. I dunno, maybe they'll last, but there will be probably a next crisis. So,

Interviewer 1

Yes. I see. Great. Thanks. And, uh, do you have any tools to help you with accessing the quality on any aspect you just mentioned?

Interviewee

Mm-hmm. . So a lot of things I tend to do myself in the, the code, but there's a lot of tools in psych learn the Python library. So most of the modeling happens in, um, in Python and they have a lot of rituals for cross validation.

Interviewee

Performance metrics, all the things I just mentioned. Um, so it'll either be that or some combination of that plus, um, my own coding of the performance metrics and under different circumstances, et cetera. Yeah. Thank

Interviewer 1

you. Have you ever used a benchmark model to, for quality aspect to evaluate your model?

Interviewee

Yes. So in, in my, in this application, the benchmark model and auto re. Model. So the baseline is kind of just what's a, what could a simple AR model do with this series? Um, it's a pretty low bar to, to cover. Um, cause it works finding normal times, but again, it's kind of useless in times of crisis. Um, so it's not hard to, to best at it, but it is at least a, a baseline where you can say, all right, this is better than the, the very simple.

Interviewee

One. And you can also use, um, the most recent, um, observation. So saying the best, uh, the baseline for the prediction is just taking the prior value. That's also a pretty low bar to, to beat, but those are kind of the two benchmarking, um, comparison points.

Interviewer 1

Okay. And do you have benchmark model for other aspect aspects than ML performance?

Interviewee

Uh, no. So those, those are good both for the, um, um, like the evaluation metric as well as the performance because they know they're fast and, um, yeah. Is it a reasonable comparison to calculation of an AR model on this?

Interviewer 1

Yes. Thanks. Yeah. Have you ever assessed a quality of a model prediction with the user of your systems?

Interviewee

So, uh, yes, actually. So the, the thing is that, Have a relatively, our, our use case as a statistical production. The goal is really to provide a resource and not, it's not the end all, be all, you know. I mean, we may have some analysis on it, but it's really to be produced as a statistic for other people to use.

Interviewee

Um, And so we don't have a lot of direct contact with users always, because it could be a student, it could be, uh, another economist at a different place. But there is the element of, um, and generally speaking, the performance metric is, is a good enough proxy for, for how useful is this thing. But there is the element.

Interviewee

Other element besides that, one of those is the,

Interviewee

The nature of these forecasts. So these forecasts change over time and they're run continually over time. So, uh, we had to kind of educate people to a certain degree. They're used to getting a forecast for the year, single number, and that's what they, you know, they run with. So for them to learn that there was, this is something that updates and changes over time was kind of a different relationship to.

Interviewee

Forecast number is for them. So there is some dialogue around what is helpful for you to see in that. And one thing that was helpful is to see the development over time. So as opposed to just giving them a single number, this is the forecast to say, this is rather presented as a line of developing over time, saying that this is the latest forecast, but this is what it was before and this is how it responded to changes in the economic outlook and in data that helps them to understand that this isn't just, you know, a number, B number and all B number, but it's one the latest one.

Interviewee

Developing trend. Um, and then there's other elements of interpretability. So people wanted to know, uh, why things changed in the model. So if I revise the forecast down by a percentage point, they want to know why, what happened? Uh, so we added that element to our presentation where it shows, um, which changes in the.

QM-explainability-fix

QM-explainability-issues

Interviewee

Resulted in that provision. It's because the German manufacturing index was published this week and it was much lower than expected. It's because the, um, the figures from last month were revised downwards because the office gotten information, which just showing people why it is doing what it's doing, and also, uh, a way to help them make sense of the tool and use it more.

QM-explainability-fix

QM-explainability-issues

Interviewee

E. Nice.

QM-explainability-issues

Interviewer 1

And you mentioned interpretable interpretability. Mm-hmm. , uh, are you using interpretable models or you're using ex explainability

Interviewee

tools? So the, the model we use is the, um, long short term memory neural network. It is not explainable in the sense of alls where you have your proficiency. And I can exactly say, oh, manufacturing portion went down by two.

QM-explainability-fix

QM-explainability-issues

Interviewee

Is 0.3. And that's exactly why you see this here. Um, they have a lot of layers. There's a lot of, there's a temporal component. It's not explainable in, um, in words really. So you have to use external, um, tools to impart that onto the model. Some degree of interpretability. Um, so this is like an exo.

QM-explainability-fix

QM-explainability-issues

Interviewee

Adjustment or add on to the model. So we use Shapley values, um, which is kinda like a, you hold out different elements of the data. So what would the forecasts have been if we didn't get the German, uh, manufacturing index last month? Compare that to what it was, and we did. And then you have some degree of, uh, of knowledge of, okay, so then this is what the German manufacturing index contributed to this, uh, development in.

QM-explainability-fix

QM-explainability-issues

Interviewee

That's, that's how we do it in Na case. Um, it's not an model native thing, and we can apply this to any model theoretically. But um, yeah, it's not an external tool. We do, uh, do it ourself. There is Aly, uh, module in Python, but um, we just code up ourselves.

QM-explainability-fix

QM-explainability-issues

Interviewer 1

I see. Thank you. And are you usually happy with the result you get with chapter value?

QM-explainability-fix

QM-explainability-issues

Interviewer 1

Is it. Good. Explain a good explainability method or

QM-explainability-fix

QM-explainability-issues

Interviewee

like the answer. So in that sense, it's like to assess the quality of the explainability is very difficult. Cause for the non networks, we don't have another way to do it. You could of course assess it, um, on os because you have the ground truth of the actual coefficient.

QM-explainability-fix

QM-explainability-issues

Interviewee

Um, but we don't really have counterfactual in the neural network. Uh, we explored other methodologies for imparting this explainability, but an important aspect of it was, um, the results. So I is this interpreter, like, like when people would ask us what is the, the contribution this gave us, the result where that's understandable to them.

QM-explainability-fix

QM-explainability-issues

Interviewee

Um, and they, they're actually was one methodology for assessing the quality of the, um, chaplain values, which was. So when you do this, you're gonna have the, um, the contributions of all the different variables. And when you add that up, ideally this should add up exactly to the change that you saw in the model.

QM-explainability-fix

QM-explainability-issues

Interviewee

And, um, so we rescale it if it's not to, to make it that. But that was one indication when we were, I was doing that process, that it was a relatively good way to do it because that number that comes out when I add up all of the contributions was almost always very close to 1.998 1.001. So that in itself is a good, uh, indicator that this is capturing why the model does what.

QM-explainability-fix

QM-explainability-issues

Interviewee

Um, like I said, it's not always exactly, sometimes it can get, you know, 1.1 or something like that, but then in the end we just rescale just to ensure that it is, does add up to the, um, total in the end. But that factor's usually very small, and that was one indication, let's say objective in indication that it's working well.

QM-explainability-fix

QM-explainability-issues

Interviewer 1

Okay. And what's the name of that technique?

QM-explainability-fix

QM-explainability-issues

Interviewee

Uh, so it's Schlafly value, like, I guess Schlafly values is the. S h a p l e y is the

Interviewer 1

approach. Okay. But what I meant is, uh, when you're accessing the quality of the chale

Interviewee

value Oh, that, uh, I don't think it has the name. It's just from my side, it was intuition that if this object was, then it's, I'm capturing everything that went on.

Interviewee

Yeah. So, yeah. I, that's a name. Yeah.

Interviewer 1

Perfect. Thank you. Um, and have you encountered any other quality issue during the evaluation of your ml?

Interviewee

No, I think we covered everything that I can right now.

Interviewer 1

Perfect. Thank you. Uh, regarding deployment, how and where are your model deployed and what, what, I mean where, I mean on the cloud or on premise?

Interviewee

Yeah. Yeah. So the, the time, two elements, there's the production of the actual model output, and then there's the hosting of the final product. Um, the final product is hosted. You know, our internal servers that's on premise. Um, but that's really just the JavaScript and, and the whatever, all the domain stuff, the actual, uh, model collecting of the data, running of the model, spitting out the outputs, happens on a server.

Interviewee

Um, that's also a non-premise server, but it's just, you know, a computer in the basement of the, of the building, um, where this will just. On a weekly basis.

Interviewee

Okay,

Interviewer 1

perfect. Thanks. And what are the challenge you have encountered during the deployment of machine learning? So,

Interviewee

tool, so the, for sure, the, the fact that this has to happen on a recurring basis isn't, I think, an underrated element. A lot of people make their model think, cool, my model, but okay, what happens next month when you wanna have your model or if you wanna scale this up to product?

QM-robustness

QM-robustness-fix

Interviewee

Level you have a lot of people will try to do things manually. I'll just run it next week, I'll run it next month. But that becomes very quickly not feasible because on a holiday you have a meeting, you just forget. Like it really had to be automatic and it had to be automatic, which meant also that it needed to be robust to problems that happen.

QM-robustness

QM-robustness-fix

Interviewee

So if, if some of the data sources don't update, I still. The result for this week, because this is a live service. I can go in later on and fix the data sources that didn't produce, but I need a, a model and a pipeline that is robust to things going wrong. So that was also an element in the model, not model selection in terms of the hyper printers, but which model we use.

QM-robustness

QM-robustness-fix

Interviewee

Can it handle a missing value? How do we handle, um, a missing value? Um, so that we have something that. Every week on the week, no problem. You can, uh, go in and fix issues if you need, if you find them. But the, the entire pipeline is robust to those types of issues.

QM-robustness

QM-robustness-fix

Interviewer 1

Okay, I see. So, so you're looking for model that are able dwindle stale data.

QM-robustness

QM-robustness-fix

Interviewee

Yeah. Okay, thanks.

Interviewer 1

Um, did you have, did you, I guess Yes. Uh, did you ever add a model that performed well locally but poorly once deployed?

Interviewee

Not in this context, no, because the deployment is pretty, the deployment is very similar to the local because it's just, you know, hosting the result essentially. It's not like people, um, kinda like a recommended system or something like that where it's, there's a lot of new interaction and individual interaction with it.

Interviewee

So, um, if you did this model inspection, Process correctly, you're not really gonna get blindsided by, oh, now the box side doesn't work anymore. Cause you, you tested on different periods, on crisis periods, normal periods. Um, you, you basically, yeah, know how it's gonna do unless, like I said, you have this mild drift.

Interviewee

But that's gonna happen slower over time. And like I said, we have processes in place. Kind of get ahead of that sense. Okay. I

Interviewer 1

see. Thanks. And did you encounter any other quality issue, uh, during the deployment phase?

Interviewee

Um, small things in terms of the, this communication between, so when, uh, So like I'm handling all the backend and i'll producing of the models, and then at a certain phase we have to pass over to our front end and our IT systems because obviously I don't have control over the Company X websites and how to deploy it finally.

friction-separated-teams

Interviewee

So there was a certain link that had to happen. Um, and in that process there was definitely some back and forth in terms of how do you want it delivered and what um means, what format do you need the data? But that was all. Smoothed out initially, some minor bug things, um, happened. All this an extra date in this, um, prediction time period.

friction-separated-teams

Interviewee

How do we handle that for the, um, for the front end, but nothing major.

friction-separated-teams

Interviewer 1

Okay, so what you're saying is you give your model to another team that deploys it on the Company X website. Yeah. And maybe the data process thing is not the same for that team as for you? Yeah. And how do you ensure that, well, you said you fix the problem, but Yeah.

Interviewer 1

How? How can you ensure them? Yeah, go.

Interviewee

I mean, for us it just happened because there's dialogue where I gave them the first draft and they say, no, we need it like this. Okay, I gave you the second draft. Okay, that's good. And I say, okay, where do you want it? How do I get it to you? Like what? Shared directory?

Interviewee

Or you want email? How do you want it? This. So yeah, just the dialogue, initial set of phase.

Interviewer 1

I see. Thanks. Uh, moving on to model maintenance, how do you ensure that the quality, well, you, you talk a little bit about it earlier on, but how do you ensure that the quality of, uh, machine learning software system does not decrease over time?

Interviewee

Yeah, I think I already explained that. I don't have anything more to say about it.

Interviewer 1

Okay. Perfect.

Interviewer 1

Yes. Okay. Alright, uh, moving on to the last section. We are going to talk about quality issues in machine learning, uh, model, sorry, sorry, . I'm still a bit sleepy. Quality measure of ML models. Uh, so we touched a little bit about it. If you have something else to say, uh, that's an opportunity. If you don't have anything, we'll just skip.

Interviewer 1

Uh, so did you ever add issue with one of the following aspect? Uh, fair. Robustness, explainability, scalability, and privacy.

Interviewee

So I think I addressed all of those except for privacy. And what was the first one you said?

Interviewer 1

Fairness. Uh, so I'm model that I have to treat everyone fairly.

Interviewee

Yeah. So that's not relevant for, for our use cases.

Interviewee

We didn't have to. Um, and then the last privacy also not relevant cause everything is publicly available. Level, national level, international level. So wasn't relevant to our use case.

Interviewer 1

Okay, perfect. Thanks. And so two, two, LA two. Last question. Uh, in your opinion, what is the most pressing quality issue researchers should try to solve?

Interviewee

Just try to solve,

Interviewer 1

it's a big question. It, it can be something that you have a lot of, yeah, a lot of issue with. And you, you say, oh, if we can solve that, life will be a lot

Interviewee

better, right? I think that it falls on the, I think there's two elements. So one is on the data collection side and one is on the model side.

Interviewee

Um, I'm generally, so I'll start with the model side. So I'm generally pretty happy with the level of like open source and resources that are available. Um, machine learning techniques and different languages, et cetera. Um, I think there's a ton of great resources, all these great libraries. The, I would just say that in specifically in the, of academia, there's too much, um, like there's enough emphasis on application and reproducibility.

QM-reproducibility

Interviewee

Not in just that, oh, I got the same results. These are, these are robust results. But in the sense of I'm now a, a different practitioner who wants to apply your methods, um, on a concrete level, I find that things are really, Abstracted, you know, they spent 45 pages on the theory model, but the next day, ok, great.

QM-reproducibility

Interviewee

I like your model, I wanna use it. And then there's nothing, you have to recode the model yourself. You have to dig, you have to, it's just not easily available. So I think write it with my paper, like this methodology is like a novel one that I developed and I made it a point that a big proponent of that is to make it a open source library so that if.

QM-reproducibility

Interviewee

I've written some papers on it. Great. You like it, you want to use it. I I may try to make it as easy as possible for, for you to do that. Um, it, this I don't think is a problem in the private sector because I think people are really good about, you know, writing medium posts. How did we do this Very, um, helpful in that community.

QM-reproducibility

Interviewee

Like specifically in the academia community. It's way too, um, siloed. Yeah, people don't share the the how they did things, not just, you know, the results. And then in the data collection side, I also think that's a super underemphasized component of ML models. People get really focused on the models themselves, but forget about all the things I mentioned earlier.

QM-reproducibility

Interviewee

All, you have to run this on a weekly basis. How do you get the data at first place? What if that one breaks? You have to transform the data first, and what if they give it to you in a different format? That whole element, I think is super under considered. Um, and there's, you know, that common saying 80% of the time is preparing the data.

Interviewee

Only 20% is actually on the modeling. Um, You know, there's no fix necessarily, like, oh, researchers should fix that. But there are kind of ways you can go, like, again, specifically in my domain, making things available via API and not just, you know, publishing an Excel file or a PDF of the data goes a long way.

Data-accessibility-issues

Interviewee

Um, and you'll see like all the data we use come from people who have invested in that process. If you're just publishing, uh, an Excel, I'm not gonna use your data cause it's too hard to incorporate into this process. So if you make things kind of easily accessible at scale and uh, on a repeated basis, that I think goes a long way in helping that data collection process.

Data-accessibility-issues

Interviewee

Um, so that's more for data providers? Um, yeah, and, and even I think like web scraping, if you. Repositories of, of pre-scripted things. I know there's a lot of people doing the same thing over and over. I need to scrape, scrape flight data or whatever. So if you can get a lot of those kind of very popular ones together into, um, standardized formats.

Data-accessibility-issues

Interviewee

I know there's a lot of private sector companies who do this, but then sell it, and that's, that's a, a barrier to entry. Maybe open source repositories of, of this flight tracker data that's aggregated so that you don't have to go in and like make your own the hundredth web scraper for this. That then breaks the next time we update the website, um, to have like a dedicated team who's like, we scrape the flight data and put it in this public repository and people can use that.

Data-accessibility-issues-fix

Interviewee

Uh, I think that would go a lot better toward. That, that's great.

Data-accessibility-issues-fix

Interviewer 1

Yeah. Thanks. And do you have any other comment about the quality of ML system?

Interviewee

I don't think so. We talked about a lot, so, and everything was covered. That's in my mind.

Interviewer 1

Perfect. Well, thanks a lot Interviewee. It was great to have you. I'm sure what you, you told us will be useful.

Interviewer 1

Uh, it'll be useful for the, the

Interviewee

art paper. So this is for your anthesis or um, nas.

Interviewer 1

Paper. Well, actually I'm writing paper, and from this paper I will build my master's thesis. Cool. So indirectly it is contributing to my master's thesis. Yeah.

Interviewee

Cool. So let me know when it's, uh, finished and published here to see the results.

Interviewee

Sure. We will. Right. Thank you. Thanks, Interviewer. Oh, really quickly, how did you guys find, like, source me, let's say like, find me to, to interview just outta curiosity. GitHub. Oh,

Interviewer 1

okay. On GitHub. Yeah.

Interviewee

So thanks guys. Good luck for your rest. Yeah, you too. Thanks.

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