

Interview 32 - Ilan

Interviewer 1

Okay. So to start off, can you give us a little bit of information about yourself? So how many years of experience you have in general and also about machine learning more specifically.

Interviewee

Yeah. Um, I graduated, uh, three years ago and, uh, immediately started working, uh, initially for, um, uh, consulting company.

Interviewee

That focus on, uh, machine learning application. I work, uh, mainly on two projects, uh, uh, that are, one is, uh, recognizing, uh, relevant info from P d F documents. Uh, could be like text, uh, inside the P D F, but also, uh, information from images that are printed inside the pdf. And the other main project was, uh, uh, classifying some emails, uh, uh, to, I mean, to.

Interviewee

Forward this emails to the right person inside the, uh, the company. And then I, after one year I moved in another, another company, my, uh, company, my actual company, uh, my current company that is Company X, uh, that is a real estate, real estate startup, uh, in Region X. And, uh, I mainly work on two applications. That one is the a v m Automatic Evaluation, uh, model, uh, that times to predict the price of a house.

Interviewee

Uh, and the other main application is, um, uh, something related to objective detection, a match classification, uh, for, from images that, uh, our sellers upload on our app and we want to give them, uh, Mm, announce their experience in the app and tell them if they missed something and they missed to upload some photos, uh, et cetera.

Interviewee

So these are the main things I worked on. Dunno if you have some questions.

Interviewer 1

No, we, we we're happy to have you here. Thank you. Uh, so I'll follow up with the really general question.

What are the main quality issues you have encountered with your data model or system?

Data-low-quality

Data-missing-values

tabular

Interviewee

So, Yeah, I think that the main quality issue is about the tabular data.

Data-low-quality

Data-missing-values

tabular

Interviewee

That we feed into our, uh, uh, a v m. So the regression model for, uh, price prediction. That's because, uh, we have, uh, a lot of missing data about some feature, but also wrong data. For example, the precise location of the house, uh, the data we have is totally inaccurate. And also the, the price, for example, that is our target variable, uh, is not precise because it's not the real price that the house is sold at the very end, but it's the price, uh, that, uh, has been requested from the, uh, owner of the house.

Data-low-quality

Data-missing-values

tabular

Interviewee

So I will say that, uh, missing and wrong data in tablet data, it's the main quality.

Data-low-quality

Data-missing-values

tabular

Interviewer 1

Okay. Uh, just to understand, the, the problem you mentioned about labeled is it that, it, just to make sure I understand correctly, is that the, the owner gives a price and you predict a price, and it's not the same, but the price predict is supposed to be the objective price and the Yeah.

Interviewer 1

The label given by the owner is not, it's subjective. It's not their real one. It's, yeah, exactly. No easy label.

Data-noisy-labels

Interviewee

So we have a lot of data with the noisy la uh, label from the owner and, uh, very few data with the actual price of the house. Cause you know, maybe we bought that house or we sold that house and we know the real price, but most of the times we don't know that.

Data-noisy-labels

Interviewer 1

Okay, I see. Great. Thank you. And, um, so you mentioned missing data. How do you address this problem?

Data-missing-value-fix

Interviewee

Um, most of the times we just dropped the, uh, the record, but, uh, for, uh, mainly when, uh, some like, uh, Uh, very important feature are missing, like the number of rooms of a house. It's, uh, very important to create the price.

Data-missing-value-fix

Interviewee

But, uh, other features like, uh, the construction era we just imputed. Um, not the, uh, the mean of the median, but, uh, reasonable value for the housing. That location. For example, we know that in City X, most of the houses are, uh, built in the 1960s and we imputed that. So not ma. Yeah, something like manual importation.

Data-missing-value-fix

Interviewee

Okay. See, so,

Data-missing-value-fix

Interviewer 1

so basically, um, your input imputing when it's not an important feature, but if it's important you're not doing it because it's too risky, fine. Yeah, exactly. Okay, great. Thank you. Um, I'll ask you a bit of question about data collection and basically I'm looking for issue you have encountered with, with certain data collection process.

Interviewer 1

Uh, so did you ever use DA data that was manually collected or manually labeled?

Data-collection-manual

Data-noisy-labels

Interviewee

Yeah. Yeah, mainly for, uh, computer vision applications. Okay.

Data-collection-manual

Data-noisy-labels

Interviewer 1

And what were some of the main quality issues you had with this data?

Data-collection-manual

Data-noisy-labels

Interviewee

Um, I'll say that, uh, if you, if I have labeled the, the datas, they don't have really, uh, quality issues.

Data-collection-manual

Data-noisy-labels

Interviewee

Uh, the, the main problem is, uh, scarcity of, uh, images because I can label just a few, uh, images for days.

Uh, but, uh, if we take, for example, a data set label by, by some other people in my company, maybe because we ask them to do that, uh, like 10 images per day, sometimes we, the, the labels are wrong or, yeah, the, the bounding box around the, uh, the object is not precise.

Data-collection-manual

Data-noisy-labels

Interviewee

Maybe it's not wrong, but it's not really precise. And this can lead. Uh, bad training results in the, in the final model.

Data-collection-manual

Data-noisy-labels

Interviewer 1

Okay. See, and for the bonding box, do you have any solution to solve this issue or you have to manually check that bonding boxes are Yeah,

Interviewee

we, we didn't look very into it, uh, with just, uh, um, starting from, uh, a pretrained model.

Interviewee

This, uh, is somehow fixed, uh, in, uh, uh, during the training step. Uh, it's if the bounding box is not totally wrong. I mean, if, if it's not, uh, very precise, uh, maybe sometimes is, uh, wrong in, uh, eh one direction, direction, uh, the next time is wrong in the other direction. And during the training phase, they did, uh,

compensate the, the wrong labels.

Interviewee

So we, it wasn't really a. Uh, big quality issue. So we didn't really look into it, uh, very deeply. I see.

Interviewer 1

See, thank you. Uh, did you ever use external data? So public dataset, third party, e p l or webs scrape

Data-collection-web-scrape

Data-missing-features

Interviewee

data? Yeah, web, uh, mainly webs scrape data.

Data-collection-web-scrape

Data-missing-features

Interviewer 1

Okay. And what were the main issues encountered for, for training with the data and also data throughout the time, like what's the reliable source of data

Data-collection-web-scrape

Data-missing-features

Interviewee

for.

Data-collection-web-scrape

Data-missing-features

Interviewee

Uh, can you repeat the, the last, uh,

Data-collection-web-scrape

Data-missing-features

Interviewer 1

question? Yeah. I'm go, sorry, I'm gonna simplify the question. Um, basically yeah. What were some of the issues you had with webs script data?

Data-collection-web-scrape

Data-missing-features

Interviewee

Oh, okay. Uh, most of the times, uh, we, we are not able to scrape all the data we need for, uh, like, uh, an, uh, listing add of the, uh, of the house.

Data-collection-web-scrape

Data-missing-features

Interviewee

Uh, so we miss some important features. Uh, for the structure we usually have in this, uh, uh, listing ads, uh, it's pretty easy to scrape data so we don't, uh, have a lot of wrong, uh, scrape data. But the, the main problem is missing data. For example, uh, something that is, uh, uh, For example, in some, um, listing platforms, they have tablet data, uh, behind the scene and they, uh, show to the user in a, in a inside the text.

Data-collection-web-scrape

Data-missing-features

Interviewee

So you should we be able to scrape the text and then analyze the text with some other models and we are not doing that, so we are missing some information because of that mainly. Okay, I see.

Data-collection-web-scrape

Data-missing-features

Interviewer 1

Perfect. Thanks. Um. and yeah. Have you ever used data that was generated by another system?

Interviewee

Mm, no. Not really. Okay, thanks.

Interviewer 1

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

Interviewee

Yeah. Uh, we try to, to measure the quality of the data by looking at, uh, uh, the variants. Inside the specific groups. For example, uh, variants of price inside the location or between, uh, properties that are really similar, for example, because they have two rooms and, uh, almost the same commercial area, et cetera.

Interviewee

So yes, we try to look into it. Uh, Um, sometimes we, for example, we drop, uh, outliers. Things that we consider outliers to improve the quality, but it's uh, just a set of rules, uh, not, uh, uh, something like very specific for the case. It's just a set of rules, for example, more than top variants. So justice group, this Rosa,

and this will lead to.

Interviewee

Improve data set actually is improved because, uh, after that the model improved the performance. So

Interviewer 1

I see what, yes. Thank you. Um, is there any other data quality should we miss that you consider relevant?

Mm,

Interviewee

no, I don't think so. Maybe yes, but, uh, , uh, I can tell it now. So. Okay.

Interviewer 1

Yeah, so if you, if you think about it, feel free to mention them.

Interviewer 1

Yeah. Thank you. Um, how do you evaluate the quality of your model? And as a reminder, quality is not only defined by the ML performance, so influence score, precision, accuracy, but also other aspects such as scalability, explainability, um, efficiency, scalability,

Interviewee

you name it. Starting from the ML per, uh, model performance, uh, uh, we usually, uh, tend to use the, uh, media and average, average error.

Interviewee

Um, but, um, sometimes it's not the best metric, uh, to evaluate our model because, uh, maybe you don't, uh, you have a low, uh, M A E A. The, the low metric, but, um, uh, one time out of two out of three, uh, your model is totally wrong. So two times is totally precise and one time is totally wrong. And this is not good for our business because, uh, if you just send the customer totally wrong number and the estimation of the price, it's just gonna drop and, uh, not gonna use our service anymore.

Interviewee

So, You just freeze. Are you able to listen to, to hear me? Uh, yes. Okay. I'm still there. Yeah. Okay. , it's just a video. Uh, yeah, so that's the main problem for us and we try to evaluate also this, um, this type of behavior, uh, with the sort of custom metric that is, uh, take the real value, uh, draw an interval of 10% around the actual, uh, the actual value.

Interviewee

The target and then, uh, how many times our predictions falls into this 10%, uh, interval. Mm-hmm. So, I don't know if it has name this metric, but, uh, this is the, the other metric we use to evaluate our ML model. And after that, uh, in terms of explainability, we also work on that. Uh, but we didn't measure what, uh, And how well the explainability perform because we integrate a, a sort of, uh, explainability layer, eh, to show to the user at the end of, uh, our valuation flow.

Interviewee

So when the user compiles a form with the extra characteristic of his house, then he receives a price range, but he also receives an output, uh, like, uh, what impact, uh, the, uh, the valuation, uh, what are the comparables, uh, et cetera. So, eh, we didn't measure the, how this impacted the, how the customers perceive our machine learning model, but we noticed, uh, that, uh, eh, the user that see an explainability layer tends to, eh, move on with our process and click on some CTAs that we have at the end of the.

Interviewer 1

Okay. You said when, when user use their explainability layer, they tend to appreciate more your product and use it in the end. Okay.

Interviewee

Yeah. For example, uh, the next step is schedule a visit and they tend to schedule more visitor, uh, if they have, uh, uh, explainability layer. I see. Makes sense.

Interviewer 1

Thank you. Um, have you ever used benchmark model for quality aspects to evaluate your models?

Interviewee

Uh, you mean for example, other a v m that tend, that try to predict the price of, uh, Anything you can compare to? Yeah, we, we tried it, but it's, uh, it wasn't really a fair comparison because, uh, uh, in the very beginning we used, uh, an external model from another company. And when we decided to switch to an internal model to build our internal, uh, uh, a v m.

Interviewee

Uh, when the job was done, we tried to compare them, but for example, we used, uh, a lot of features and they just use, uh, like five feature to predict, uh, the model price. So we, we knew that, uh, our model was better for sure, because they just use five features, but it was done because, uh, to convince business people to adopt our model and.

Interviewee

Yeah, that's, that was the main comparison we did. I see. And

Interviewer 1

was there some issue, you, you, you have found once you compare the model issues in your model or, uh, not

Interviewee

really. Um, no, not really, because what, uh, as I said, the other model are just five features that he perform worst on most of the records, not just on, uh,

Interviewer 1

The point was to show that your model was better.

Interviewer 1

So, yeah. Hmm. Yeah. Thank you. Um, have you ever assessed the quality of your model prediction with the user of the

feedback-loop

Interviewee

system? Yeah, because, uh, we was also this model, uh, eh, for internal purposes, uh, Like, uh, if, uh, if we see like our business person that, uh, goes to the listing platform and see an interesting house, they tend to evaluate this, uh, this house also with our model.

feedback-loop

ungeneralizable-problems

Interviewee

And, uh, they use it as a benchmarker. They start from the asking price of the other. They use our model. They do a manual evaluation somehow. And they see if that they, they can be compared, they earn the same magnitude. And, um, if it's like that, uh, they goes on without any problem. Uh, if they see that their evaluation, evaluation, it's very different, uh, from the hour evaluation, they, maybe they talk with other colleagues to check, uh, if they did something wrong, et cetera.

feedback-loop

ungeneralizable-problems

Interviewee

But the feedback that, uh, arrived to us is that, uh, mm, like half of the, of the time, it's very accurate. Half of the time it is not reliable at all. So it really depends on the zone and on the eristics of the house. So for some houses, very precise. For some other types, it's not precise at all. I see.

feedback-loop

ungeneralizable-problems

Interviewer 1

Okay.

Interviewer 1

So basically the model is not a, is. doesn't generalize to all the distribution basically.

Interviewee

Yeah.

Interviewer 1

Thank you. Um, is there any other quality issue about the evaluation of machine learning software system that we miss, that you consider relevant?

MM-drift

Interviewee

Uh, yeah, that's a, uh, huge problem we are having now. It is that, uh, the price of the house evolves in, in time.

MM-drift

Interviewee

Uh, so if we want to, for example, update the comparables we use inside our model. So the other process is we com uh, we do a comparison with the price for, uh, give the prediction. Um, we also have to evaluate, uh, how, uh, this new model is better or worse from the other. But the problem is that we cannot keep, uh, the old comparables for a year, for example, because from a year to another, the price may, may change for like, uh, a lot, 10%, for example, uh, for the inflection, uh, et cetera.

MM-drift

Interviewee

So we tend to just promote the production, uh, almost, uh, every update, uh, in terms of, uh, data, data, not, uh, model structure of data. But, uh, sometimes we are not very sure about it because maybe the data quality lowers somehow and we didn't realize it. And, uh, we promote a production, uh, model that is worth the previous one.

MM-drift

Interviewee

So this is one of the quality issue we are, uh, facing now.

MM-drift

Interviewer 1

I see. Okay. So just to be clear, so you have a model and you create a new model to, to address the concept shift in your data and the new model. You cannot compare it to the old model because the old model is already outdated, so you have nothing to compare.

Interviewer 1

More or less. Okay. And, and so one of the issue of updating, so frequently your model is data quality issue, as you mentioned. Uh, do you have some tool to detect that there is problems in your data set and also to fix them or, it's pretty

Interviewee

much manual? Yeah, it's pretty much manual. Uh, we have some, uh, like automatic queries that tells us, for example, Um, they asking closing spreads, we can check if they are reasonable or not, and some other like, quality queries.

Interviewee

We can call it that way. But, uh, the process to check this, uh, numbers and promote production, a new model is totally manual. Okay, I see.

Interviewer 1

Perfect. Thank you. Um, all right, so moving on to deployment. How and where are your models deployed?

Interviewee

Um, we usually, um, like, uh, Python scripts, uh, with a fast api. Inter, uh, interface.

Interviewee

Uh, we have a docker containers. Uh, we have, um, a continuous integration pipeline that build the docker container, uh, in, with the first P up and deploys it with the helm on a Kubernetes cluster. But, um, this is, uh, may already written by someone so I don't have to like, Change lot of things every time. I just like push on and, uh, wait to the model to be deployed and maybe once every six months, uh, check, uh, like change some configurations.

Interviewee

Uh, but uh, yeah, it's not my main job.

Interviewer 1

Yes. I see you're a data centers. Yeah. Yeah, exactly. Yeah. Um, and where, where is it deployed? Is it on the cloud or on premise?

Interviewee

Yeah. Yeah, on, uh, e k s. It's the name of the Kubernetes cluster for aws, I think. Ah, okay.

Interviewer 1

Yeah. Perfect. Thank you. Um, what are some of the, of the challenge you have encountered during the deployment of machine learning software system?

MLP-immature-GPU-services

QM-efficiency

Interviewee

Um, Not specifically for this use case, but for the other use cases, uh, where I use, uh, like the learning models more, bigger models, uh, uh, one of the problem is like, uh, uh, call starts, uh, because you have to, uh, upload memory your model. And, uh, it takes, it may take a lot of time. Uh, Not a lot of time up. 10, 10 seconds if you, if you don't have a pod already there for, uh, receiving a request, the user to wait 10, 10 seconds.

MLP-immature-GPU-services

QM-efficiency

Interviewee

It's not really great. Um, the other problem, uh, is, uh, like using GPUs because, uh, we have, uh, we are not using in, uh, Company X this type of technology, but, uh, serverless functions, for example, you can easily use them with the cpu. But, uh, when we try to deploy a model with gpu, we also consider that, uh, that path.

MLP-immature-GPU-services

Interviewee

But, uh, we discard it because, uh, the, the tools are not very mature to serverless GPUs because, uh, they don't, uh, uh, handle very well the, the cost starts and, uh, the, and other things about, uh, about it. No, I don't know if it's clear or not.

MLP-immature-GPU-services

Interviewer 1

Uh, more or less, I will tell you what I understand. So, so one of the issue you have is CodeStar.

Interviewer 1

So, so when you, when a model is already deployed, uh, the first query may take more time than the other query. Right. Okay. And then you mentioned CPU and GPUs. Mm-hmm. , and this is the part, maybe I misunderstood. Uh, when you talk about

Interviewee

this, Yeah, the problem is that, uh, eh, it's hard to work with GPUs on, uh, using Kubernetes clusters or in serverless, uh, tools.

Interviewee

Because there, I think that, uh, it's not really mature or maybe, eh, also infrastructure guys are not very, uh, they don't face a lot of, uh, issues, uh, in the past with the GPUs, uh, et cetera. So also the experience is lower in the market for this type of, uh, things. So yeah, that's the main problem. So at now we are just using, uh, CPUs.

Interviewee

For our models and we try to make the model smaller, maybe losing some, uh, performance, uh, eh, but yeah, at least it can, uh, run in, uh, like one second and not more. Yeah. Yeah.

Interviewer 1

Okay. That's fair. Thank you. So, and so the issue is, um, server serverless architectures, well, serverless architecture that are. given by aws.

Interviewer 1

Uh, they're not major in term of GP of using GPUs. It's, it's not, it's a bit complicated. Okay. Okay, perfect. And is, is there a, a link between the COLDSTAR and the GPUs? Like it's more difficult to have a outstart with GPUs or It's two different problems.

Interviewee

I think it's, uh, two different problems. Uh, uh, I mean, they're somehow correlated because, uh, if you want to use the GPU user, maybe you can avoid the problem of call starts because, uh, it costs less to have, uh, a machine already, uh, always running with the A C P U.

Interviewee

But if you want to do it with g p. It will cost a lot to, uh, have a machine always running up and running. So maybe you want to switch to a serverless approach and, uh, it's hard to do it. Yeah. So maybe there. Great.

Interviewer 1

Perfect. Thanks a lot. Um, did you ever have a model that per perform well locally but poorly once deploy?

Interviewer 1

I think you al kind of already mentioned that, but if

Interviewee

you want to a model that perform, uh, poorly. In production and well, yes. Yes. Um,

Interviewee

well, I'm not sure about it. I mean, yes, because, uh, eh, when we try to analyze, uh, the, uh, the performance of a model, uh, built in the past for the current data, This may perform well with our test set. Uh, that is built always in the past, locally, but when we are using production, it doesn't perform very well.

Interviewee

So I think that's the, the only, uh, the only occasion where this, uh, course, but, uh, for other application, like computer business application, uh, it, uh, usually doesn't happen because, uh, if it, if you build a. A good test set locally, it'll perform almost the same, uh, also in production, uh, because images doesn't, are not affected by time shift or something like that.

Interviewer 1

I see. Thank you. Um, and is, is there any other quality issue during the deployment that we miss and you'd like to mention?

MD-versioning-issues

Interviewee

Um,

MD-versioning-issues

Interviewee

Not really May, maybe how you, um, version the models and how you retrieve the model, uh, when you deployed it. Like, for example, uh, in the previous uh, company I was in, uh, we just, uh, put the last version of the model on s3. And then we, uh, when you start the, started the application, uh, the model was fetched with the, uh, bottle three, uh, library and, um, instant in, uh, in ram.

MD-versioning-issues

Interviewee

But, um, but now we are using D vc. And during the GI Pipeliner, we just fetched that, uh, the data with dvc and, uh, we package the model inside the, the container. So, and I'm not really sure what, uh, what's the best approach. So it's not, uh, really a, a standard. There's not really a standard to do it. And uh, yeah, every time you just try to build a new machine learning system, you have to face also this issue, how to fix the product production.

MD-versioning-issues

Interviewer 1

So, so basically you mentioned two ways of automatic deploy of automatically deploying model. Yeah. The first one is you build model. Following per periodically. Let's say you apply to S3 and then, and then, sorry, I'm not sure what follows

Interviewee

up. Yes. You patch the model when you start the application directly.

Interviewee

Okay.

Interviewer 1

Yeah. Okay. And the other way around is you don't build the model periodically. You only do it when once you deploy and when you deploy. It's a data trained creator computer. I see. Okay. Okay, perfect. Thank. You're

Interviewee

welcome, Adam. That was it.

Interviewer 1

Yeah. Yeah. Well, I, I, I was thinking of a follow up question, but right now it's not coming up.

Interviewer 1

So Yeah. , sorry, Frank. Okay. Um, how do you ensure that the quality of machine learning software system does not decrease over time? Um,

Interviewee

yeah, we. As I said, also internal users, uh, for, uh, coming back to our, uh, a v m application. Uh, internal users, uh, uh, tend to use this model, uh, to, uh, evaluate properties. Uh, and then they also sign a manual, uh, value to the, to the property price.

Interviewee

And, uh, We, we just compare these two information we have. So the prediction of the model and then manual priced from the, our acquisition specialist. And, uh, we see that, uh, eh, the difference in terms of, uh, medium, square, eh, average error, an absolute error, sorry. Um, and it usually, For like three, four months.

Interviewee

Uh, the performance, uh, eh, don't degrade over time. Um, so yeah, and as we update the comparables like every two months, uh, three months, uh, it's not really an issue for now for us. Perfect. Thank you.

Interviewer 1

Um, have you encountered, sorry, we already covered. Uh, have you encountered issue with the model during the maintenance of machine learning software system?

Interviewee

Sorry,

Interviewer 1

can you repeat? Ha. Have you encountered issues with the model during the maintenance of machine, machine learning software system?

Interviewee

Mm, I'm not sure what you're referring about. I mean, what do you mean for maintenance? Of the machine learning, uh, well, we

Interviewer 1

already talked, but one example is models.

Interviewer 1

Models stay or less. Uh, sometime also people tell, tells us that, uh, the, the retraining of the model is unreliable between each deployment you already mentioned a bit also.

Interviewee

Yeah. And I don't know if the, there are other issue issues, uh, I faced in my experience because, uh, uh, Uh, this is the, the a om is the only model, uh, I, uh, I did maintenance on in my experience because, uh, for example, the computer vision models we have in production, uh, we released them in, uh, like one year ago and we did, uh, never touch them because, uh, they're performing well also now.

Interviewee

So, yeah, I don't know if there are other issues I have to make. Okay. That's perfect.

Interviewer 1

You already mentioned. . Um, so for the next question, I'm gonna list a few quality aspects, and basically you tell me if you, you, if you have any, ever had an issue with one of them, please share, share me the, like, the, your experience with it, right?

Interviewer 1

Mm-hmm. , uh, so fairness, robustness, explainability, scalability, data privacy, and model

Interviewee

security. Uh, Okay. Can, can we go one by one?

Interviewer 1

Yes. Sure. Um, so fairness, did you ever add issue with

Interviewee

fairness? Yeah, for fairness? You mean? Uh, like, uh, how accurate is it? Or,

Interviewer 1

or no, it's, um, let's say you have, um, sensitive group.

Interviewer 1

Okay.

Interviewee

Let's say women and men answering. Uh, okay. In terms of, okay. No, no, I didn't never face this kind of. ,
okay.

Interviewer 1

Uh, robustness.

Interviewee

Yes. If, uh, for example, the program, I mentioned that, uh, for some types of, uh, groups, uh, the, the a m
performs well for other types. No. So, yeah. Yeah. It's kind of robustness, right?

Interviewee

Mm-hmm. ,

Interviewer 1

uh, explainability.

Interviewee

Yeah. Uh, we, I already talked about how we managed, managed the explainability layer, but to build there, there were a lot of issues we, we have to face. For example, uh, what explainability, what, what is the explainability we want to show to the user, uh, how to. Uh, explain also the explainability part.

Interviewee

For example, if we show the comparables, uh, what, uh, we can say, okay, the model is based on these, uh, three comparables properties, but what is a comparable property? So for our perspective is clear. Maybe for the customer perspective is not so, yeah.

Interviewer 1

Can you give me an example of comparable proper?

Interviewee

Yeah, for us is, uh, just a, a property that, uh, we already saw the real price or the expected price.

Interviewee

And, uh, we see that, uh, it has a, uh, really, they're, it's, uh, really like your property. They, they have similar characteristics to your property and, uh, we just use, uh, their prices to adjust the machine learning. So for examples, if the machine learning price is 1000, but all the comparables are 1,200, maybe the final price is, uh, average of the two type, uh, this two type of system.

QM-explainability-issues

Interviewee

Okay? But we, we have to explain this type of things to the user, to the final user to gain this trust. And it's not easy . Okay?

QM-explainability-issues

Interviewer 1

And why is not, is it, is it not easy?

QM-explainability-issues

Interviewee

Cause it's a little bit technical. And, uh, if you want to explain our model, uh, is built to a non-technical person, sometimes could be hard. Also, explainability tech techniques that are really famous, like, uh, eh, shop values, for example.

QM-explainability-issues

Interviewee

Okay. Eh, they're really helpful for technical people in my opinion, but, uh, they can be hard to inter be interpreted by non-technical people. Okay. See, that's a really

QM-explainability-issues

Interviewer 1

good point. Thank you. And so basically, instead of using these complicated, um, measure that are really technical, uh, you do some of your own, some expon metal of your own, but it's kind of difficult to to build them.

QM-explainability-issues

Interviewee

Yeah, exactly. Okay,

Interviewer 1

thanks. Um, and the next point was scalability.

QM-scalability-issues

Interviewee

Yeah. Uh, we faced, uh, scalability issues, uh, Uh, for example, at now we have, uh, one model for each city we operate in. So there are like nine models plus other three models, one per country. And, uh, they're all the, these models. Uh, Are in the same pod, Kubernetes pod.

QM-scalability-issues

Interviewee

And, uh, we need to have like a lot of, uh, memory located, uh, for that. Uh, and we start planning to split the models, uh, in different pods. But, but then you have to, uh, address the, the requested to the models, what the right model. Uh, so for example, if arrives a request in City X, you have to build an intermediate player that enters the risk requester and others the request to the pod with the model from City X.

QM-scalability-issues

ungeneralizable-problems

Interviewee

So a lot of scalability issues in, uh, having lot of models for the same purpose.

QM-scalability-issues

Interviewer 1

I see. So basically because a model cannot generalize to any distribution or to the city, you have to do one per city. And this is really tic to, to build the infrastructure. Thanks. Uh, and the next point was privacy, data

Interviewee

privacy.

Interviewee

Um, no, I never, I mean, mm, usually I don't, uh, use, uh, uh, data that are quite sensitive, uh, because for example, yeah, the, the house details are not really sensitive. You don't have the email or the name of the owner, and we don't use this kind of data, but, uh, We are facing some privacy issues, uh, in the, in the feature we want to build it now, uh, that is share between users, some photos for examples, and, uh, for example, there could be a photo where there's a person inside it and maybe we want to, uh, just drop the pictures or just, uh, do a segmentation of the persons and remove it from the imager, but, Just, uh, privacy issues to the whole application, but not on, not only data because, uh, eh, we can treat this kind of data because we have authorization from the users to treat this kind of data.

Interviewee

Okay. And

Interviewer 1

the, the follows from the user, uh, they're not used to training any model. It's, it's more to, it is to show pictures of the place. Yeah. Basically. Yeah. Thank you. And the last point is, uh, security of your models. Sorry, uh, sorry. The last point is security of your models. I

Interviewee

can't hear you.

Interviewee

The noise. It's, it's my fault. Or,

Interviewer 1

uh, can you hear me now? Yeah, yeah. Okay. Yeah. Perfect. So the last point is security of

Interviewee

your models. Okay. What do you mean for security? Uh,

Interviewer 1

for example, uh, you don't want your model to be stolen. This

QM-model-security

QM-model-security-fix

Interviewee

is one example. Oh, okay. Yeah. Yeah, we, we face this issue because, uh, the, like, uh, when you try to evaluate your house, uh, you have like, uh, uh, a website page when, uh, where you can just fill the form in different steps and then click on evaluator, and this triggers the model and it gives you a.

QM-model-security

QM-model-security-fix

Interviewee

But, uh, the problem is maybe that, uh, a bot can do a lot of request and try to just evaluate every kind of, uh, house there's on the market. And, uh, they can use as, uh, like, uh, uh, teacher model to train another model. Uh, and I didn't work on that, but I know that the team, uh, work hard on that. And, uh, they like introduce a sort of, uh, recapture.

QM-model-security

QM-model-security-fix

Interviewee

To avoid the bots, to do a lot of request to the model. Oh, that's really

QM-model-security

QM-model-security-fix

Interviewer 1

interesting. Thanks. Uh, I have two questions, but just before that, I'd like to go back, uh, on what you, you, me, something you mentioned earlier and I didn't do a follow up. Uh, I think you mentioned some versioning issues. Well, it, it was when you, you mentioned how the models were deployed, basically.

Interviewer 1

Yeah. Uh, would you go, would you mind, did you ever have versioning, versioning

Interviewee

issues? Uh, yeah, I mean, with the last approach, we are using it now in the, um, my company, we're not having, uh, versioning issues because every time we train, uh, a new model, we just version it, uh, uh, on s3 and, uh, it's, um, It is also tracked with the code on, uh, on GitHub, because with d c, you just linked the, maybe you, you already know this, uh, uh, this library, but mainly you link the GI dash to, uh, a set of data.

Interviewee

You have, you can just fetch them with the, like from three in our case. But yeah, in the past, uh, yes because, uh, we just pushed on s3, the last model. Because we didn't have the infrastructure to version models, uh, and to version performance of the models. So maybe it was also hard to know what were the performance on the test set, uh, from the model that was in production, for example.

Interviewee

Yeah. That's a version in issue. Okay, I see. Thank

Interviewer 1

you. Uh, and yeah, to finish, um, in your opinion, what is the most pressing quality issue? Researchers should try to.

Interviewee

Huh, . That's a tough question. Um, mainly, I dunno, because they're researchers maybe don't, don't have, uh, access to like, uh, eh, all of the data companies have access to.

Interviewee

So, and maybe also scaling issues, uh, are hard to address, uh, in the research phase. So maybe the, the evaluation. Like, uh, evaluating, uh, uh, not, not just the metrics that, uh, should be used to evaluate a model, the performance of a model, but also how to compare models, uh, be, uh, between time and, uh, with, uh, also compare models, uh, that are trained on different type of data, maybe.

Interviewer 1

Okay. What, what do you mean by the models that are trained on different types of data?

comparing-models-different-features-difficult

Interviewee

Yeah. For example, uh, let's say that, uh, during, uh, the, the training phase you have access, uh, to a specific features feature. And then, uh, after one month you don't have access to that feature anymore because like, uh, your data provider doesn't provide you that.

comparing-models-different-features-difficult

Interviewee

Uh, and you have to compare the two models, but maybe the last, uh, the second one we perform always, uh, lower than the previous one because that feature was important and maybe you have to try to compare this, uh, this models for, to the decide whether to push in production, the model or not. And yeah, this is a feature, this is, this is an issue that maybe can be others also by researchers like you.

comparing-models-different-features-difficult

Interviewer 1

Yeah. So basically it's to fairly compare models, train on the same problem, but slightly different feature set.

Interviewee

Yeah. Features or, uh, like, uh, uh, data, uh, from different times, uh, time ranges. Okay, I see.

Interviewer 1

Oh, that, that's, yeah, that's a great point. Thanks. And to finish, do you have any other comment about the quality of machine learning software?

Interviewee

Mm, not really. We really talk about a lot of things. . Yeah,

Interviewer 1

just, just to make sure. Anyway, so thanks a lot Interviewee. I think it was really interesting and you shared with us a lot of in important information that will be useful for our study. So we're really grateful for your time.

Interviewee

Thank you. And will you, uh, just for example, not, uh, let me know when you publish the paper cause I'm really interested in Yeah, sure.

Interviewee

Sure. Okay. Thank you. Thank you for your

Interviewer 1

interest. Yeah. Anyway, so thanks again and have a good day. You too. Have a good day. Bye.

Interviewee

Bye. C.

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