Transcript

00:00:00 Interviewer

Yeah, so I'm recording this session after getting the permission from you. So I think you have questionnaires open in front of you or otherwise I can share my screen.

00:00:10 Interviewee

No, I have them open. Yep.

00:00:12 Interviewer

OK, so if we go to the first questions, can you introduce yourself and describe your job role in the current company or in the current university or?

00:00:22 Interviewee

Sure. So I'm “ ”, maybe a little out of context, but I just switched to a new company just this week.

00:00:32 Interviewee

But previously I was a senior manager for data engineering for one of the multinational investment banking companies currently based in Sydney, Australia.

00:00:42 Interviewee

Better my previous company, we operate in different countries in Houston, in London. So my focus is mostly about dot engineering. So, it's it was all about building end to end data pipelines.

00:01:00 Interviewee

Getting data from various source systems, doing the consolidation integration and making them available for different business units for consumption, such as for analytics and machine learning purposes.

00:01:13 Interviewer

Oh, nice, yeah.

00:01:15 Interviewer

So, since how many years you are working in this company like your previous company?

00:01:20 Interviewee

So, for my.

00:01:21 Interviewee

Yeah, it was just last week, so it's fun. I've worked for them for almost one year.

00:01:26 Interviewer

OK. One year. So total, how many years of experience do you have?

00:01:30 Interviewee

I currently have about 15 years of experience, all of those being in the data analytics and machine learning industry.

00:01:33 Interviewer

OK.

00:01:39 Interviewer

OK, nice. So have you published any thesis in machine learning?

00:01:44 Interviewee

No, I'm uh. I'm actually on the extreme side of being a practitioner. So uh, most of my experiences, really.

00:01:49 Interviewer

OK.

00:01:53 Interviewee

You can call it like applied kind of machine learning applied analytics because they're mostly in the context of actual industry use cases.

00:02:03 Interviewer

Oh. Oh, nice.

00:02:05 Interviewer

So can you share your experience in your current position?

00:02:11 Interviewee

Yeah. So I guess you would like to hear more specifically about machine learning or in general?

00:02:16 Interviewer

Yeah. Yeah. Specifically the machine learning, yeah.

00:02:19 Interviewee

Yeah, sure. Sorry. Specific to to the current company, right, not with the previous ones.

00:02:26 Interviewer

It can be the previous one because the one week how much you can learn.

00:02:32 Interviewee

Uh, yeah. I mean, maybe I can. I can. I can talk to you in general because.

00:02:36 Interviewer

Yeah, that would.

00:02:37 Interviewer

Be really cool.

00:02:38 Interviewee

Yeah, I have.

00:02:38 Interviewee

I did have various experiences in machine learning in different companies. So yeah, my experience in machine learning I think could practically.

00:02:47 Interviewee

Say maybe about four to five different use cases, all of which are in the banking.

00:02:53 Interviewee

Industry. So First off, I started with the traditional statistical modelling requirement which is around the credit risk analytics.

00:02:54 Interviewer

OK.

00:03:07 Interviewee

I used to work with a consulting company and the main task was to try to optimise the existing machine learning model that they have. Well, the existing statistical model that they have.

00:03:23 Interviewee

There was a that that bank actually hired a third-party credit rating company to help them out with the statistical model, and one of the main things that was recommended to them was the use of genetic algorithms.

00:03:38 Interviewer

OK.

00:03:39 Interviewee

I mean, I'm going to talk in general about modelling, machine learning and statistics, you know? Yeah. So.

00:03:46 Interviewee

Coming in as a consultant, we were asked to kind of do the implementation of the genetic algorithms recommendation that was provided to them as a consultant.

00:04:02 Interviewee

It is very typical for us to deliver beyond the requirements, most especially when we feel that.

00:04:06 Interviewer

Yeah, exactly.

00:04:11 Interviewee

We can do so much better than the requirements. So simply put, we were able to provide to them the technical implementation of the genetic algorithm.

00:04:22 Interviewee

But on top of that, since we add, we want to add more value. We kind of enlightened the customer that.

00:04:31 Interviewee

It's not necessarily about the machine learning algorithm or technique.

00:04:39 Interviewee

A lot of a lot of the optimization can actually be done just with the.

00:04:45 Interviewee

Proper data cleansing, data integration and proper. Proper choice of variables.

00:04:52 Interviewee

So, I remember there was about 50 variables that they were looking at initially and then we just, you know we just followed it by the book. OK, let's use 50 variables. Let's use all the existing scoring that you.

00:05:04 Interviewee

We have. Yeah. Let's just do it. And I think we gained about just the 40% model accuracy.

00:05:12 Interviewee

And then after that, we told them, hey, we can do so much better and it's not necessarily, you know, it doesn't have specifically to be genetic algorithms.

00:05:21 Interviewee

So, we started guiding the customer and showed to them. First off, do the proper data cleansing trimmed down these 50 variables. I think we trimmed it up. We trimmed it down to about 20 or even 16.

00:05:34

I think.

00:05:36 Interviewee

And then from there.

00:05:38 Interviewee

That's where we show.

00:05:38 Interviewee

Them then, when your foundation of data cleansing foundation of having the right variables in in place.

00:05:46 Interviewee

That's where you can freely choose which algorithms you, you would want, because now you know your base data is solid, your base data is is really correct and from there that's when we show, when, where, when we show them. OK you know we can try using basic regression we we use native based we used.

00:06:06 Interviewee

Random forest. I think we also tried SVM. I mean we tried. We tried various different kinds of algorithms.

00:06:16 Interviewee

And truth be told, it was still random forest who yielded the most accurate result.

00:06:23 Interviewee

Right. But I think I think this is where this is where academia versus industry kind of comes in.

00:06:31 Interviewee

Because black box algorithms it's very hard.

00:06:36 Interviewee

To explain it to to the business, to the business like hey, so why are these variables, you know stronger than the?

00:06:42 Interviewee

Other ones, so I mean.

00:06:45 Interviewee

Can you explain why? I mean, why did the models you know say it like that and during those discussions we, the ultimate result was the customer shows the regression model just the simplest model that you can.

00:07:01 Interviewee

Have, but it was actually the most fundamental.

00:07:04 Interviewee

Is that model you can explain each variable each way you can explain it straight to the business and those learnings, those insights that they have can actually be fed back to the to the front end of people, the one dealing with the customers themselves. So kind of kind of detailed it out.

00:07:25 Interviewee

In a very specific way, but at a going back at a high level, I think there are two or three important things that I've mentioned in there #1.

00:07:35 Interviewee

Before you dive deep into the algorithms, the most important thing is making sure that your dot and your variables are very are very good, very accurate, very good, very clean. So in spite of whatever algorithms you implement, if this foundational thing is problematic.

00:07:54 Interviewee

You're not going to get anywhere #2.

00:07:59 Interviewee

Being able to try out various different machine learning algorithms is what do you say?

00:08:07 Interviewee

It's very valuable as well because it gives it shows that you're flexible, it shows that you're flexible enough to try what whatever is, whatever is needed. But but Third Point.

00:08:18 Interviewer

OK. Yeah.

00:08:22 Interviewee

It's not always about the most optimal model.

00:08:26 Interviewee

It's actually what is the most useful for the business.

00:08:31 Interviewee

Bit between I think we we did that about four or five years ago between that time and this time I think black box algorithms have become more explainable.

00:08:42 Interviewee

But still when you talk with business, I mean when you get into the industry being able to communicate that to the business is still the most essential part.

00:08:53 Interviewer

Yeah, that is very well explained, yeah.

00:08:56 Interviewer

So towards the 5th question, do you have any experience in the previous company which is developing machine learning system? I think you have already explained.

00:09:05 Interviewee

Yeah, I mean that was one example I have of quite a few more but let me know if you if you want to hear some more. But yeah, that was the.

00:09:10

It.

00:09:11 Interviewer

It will be fine if we can hear one.

00:09:14 Interviewee

Yeah, sure. I think the, the, the next couple, I can wrap them as one because the next use case was about natural language processing. So again, these are for two banking clients and.

00:09:34 Interviewee

Uh, yeah, let me explain. So the first one was, uh, it was more about an operational risk requirement where we were trying to, we're trying to catch employees who are having sales small practises.

00:09:51 Interviewee

So it was more of an internal audit, internal risk detection kind of initiative.

00:09:59 Interviewee

And we were doing that using e-mail data.

00:10:06 Interviewee

So again, uh, with uh, an unsupervised we the the main challenge for that one is we didn't have any labels of what is a a bad e-mail or a fraudulent e-mail. So the only technique that we can that I can initially do was.

00:10:26 Interviewee

An unsupervised learning.

00:10:28 Interviewee

Uhm, so I was given this huge e-mail data set with no labels in it. All I can do is to apply unsupervised learning techniques, various clustering techniques, blah blah.

00:10:43 Interviewee

By the end.

00:10:44 Interviewee

Of it I came up with I did some topic modelling as well. By the end of it.

00:10:48 Interviewee

I came up with some themes of what these emails look like.

00:10:53 Interviewee

And I would say that.

00:10:56 Interviewee

The May.

00:10:58 Interviewee

After having prepared that data, for that to be usable for the business.

00:11:05 Interviewee

You would still need to undergo that manual process of people verifying of whether the e-mail.

00:11:12 Interviewee

Is indeed a fraudulent e-mail or not. So if you zoom out, that basically means.

00:11:21 Interviewee

Doing the labelling on the data.

00:11:24 Interviewee

And again, this was about three years ago, I think for the past couple of years, there's been a lot of startups who has launched services to do labelling services. But in spite of that, if you work with huge enterprise companies such as banks.

00:11:44 Interviewee

This daughter.

00:11:45 Interviewee

They can't just.

00:11:46 Interviewee

Simply go out of the go outside of the.

00:11:49 Interviewee

Company. So. So those labelling services, typically you need to engage them, you need to send out the all the data sets so that ain't that's not really possible. So what what's what? What happened was we needed to engage.

00:12:05 Interviewee

An operational team, an operational business unit where I gave them all these themes and emails and they analysed manually each one of them.

00:12:16 Interviewee

They went through each one of them and they fed me back to what they have verified as fraudulent emails or not.

00:12:23 Interviewee

So from that, you know, so it's kind of.

00:12:26 Interviewee

I helped you categorise these emails.

00:12:28 Interviewee

Of what I think is, are.

00:12:30 Interviewee

Bad are bad. Emails are fraudulent emails.

00:12:33 Interviewee

As much as possible to some.

00:12:35 Interviewee

Extent of what I can do.

00:12:37 Interviewee

You know applying NLP techniques. I mean I'm not going to go specifics but applying various different NLP techniques, creating my own dictionaries.

00:12:48 Interviewee

And then gave this to the operational business unit.

00:12:51 Interviewee

They did the labelling from there. I add this small sample data set of labelled data. Then I was able to introduce a hybrid approach in in the model.

00:13:02 Interviewee

So it's not just unsupervised learning technique anymore. I was able to do a hybrid with some supervised learning techniques and I was I was able to kind of.

00:13:11 Interviewee

Do an ensemble approach to come up with a like a fraud detection kind of e-mail.

00:13:21 Interviewee

But again, just to give you some some practical perspective.

00:13:26 Interviewee

Applying machine learning in huge companies, it's not like how do.

00:13:34 Interviewee

I say it.

00:13:36 Interviewee

You don't get to apply it like ohh I came up with some fraudulent emails, they verified it. I get some accuracy.

00:13:42 Interviewee

It's not like you can't put it in production. You can't. You can't operationally operationalize it right away like like real time detection is of some sort, you.

00:13:51 Interviewee

You don't because.

00:13:52 Interviewee

Why? Because it's a huge company. You don't want typically huge companies don't like undetermined tic programmes to be to be deployed in real time, so typically.

00:14:12 Interviewee

They deploy rule-based approach in real time and then the model based approach is more like in the background as a detection mechanism. So you know I'm just saying that.

00:14:26 Interviewee

Machine learning like complex or maybe machine learning algorithms are typically being deployed especially for risk. Use cases are typically being deployed at the back end, some sort of a detection mechanism, but what gets deployed on the front-end real time or actually rule based rule based approaches.

00:14:45 Interviewer

OK. Yeah.

00:14:47 Interviewee

Yeah, I think that that's one key thing to understand. I think from an academic perspective versus the what's in the industry because sometimes you know, ohm I have an 80% accurate model just deployed in production.

00:14:59 Interviewee

You know, have it, have it. Have it detected fraudulent emails in real time. But no, companies won't do that. They would still want.

00:15:00 Interviewer

I see.

00:15:06 Interviewee

I mean it's more assistive, like OK, if there's some manual people trying to detect what fraudulent emails are and on the side they will see, oh.

00:15:14 Interviewee

The machine learning model detected that this is a fraudulent e-mail. You know it's more assistive, it's more to help people who are doing the manual stuff make things easier for them, yeah.

00:15:25 Interviewer

Exactly. Yeah, it is.

00:15:29 Interviewer

So, is your previous company is service based or product based?

00:15:34 Interviewee

So, I've mostly worked with the service-based companies I worked. I mostly worked with consulting, but other than consulting I was working with the various banks so it it was a mix of a service and.

00:15:46 Interviewee

Product based, yeah.

00:15:49 Interviewer

So what software development model do you practise in in your company? In journal like Agile or Waterfall?

00:15:56 Interviewee

I've been in industry for 15 years. The first half of it, mostly people were doing waterfall. But starting 2010, 2012.

00:16:05 Interviewee

Everybody has moved to Agile, so it's very seldom that you see people doing waterfall approaches these days.

00:16:13 Interviewer

OK.

00:16:15 Interviewer

So could you place your experience with interesting projects in machine learning that you have worked on recently?

00:16:22 Interviewee

I'll keep this brief this time, so it was the other NLP project we're in. It was more like a information extraction.

00:16:30 Interviewee

We were given a bunch of PDF's like like PDF scanned PDFs, so we had to do some OCR approaches.

00:16:41 Interviewee

Text detection identification techniques so that we can specifically extract information from those documents, such as financial information, customer information and build a timeline of the interaction of those customers with the company.

00:17:01 Interviewee

So it was very interesting because.

00:17:05 Interviewee

Like I told you, like heaps of people were doing it manually. I think it was about 50 to 100 people were doing it, but we prototyped the the the NLP programme that we have and we were able to achieve better results than than those people, yeah.

00:17:10 Interviewer

Yes, it helps them.

00:17:25 Interviewer

Yeah. Thanks.

00:17:26 Interviewer

So in your working experience, how many software architecture, design techniques of machine learning your work?

00:17:34 Interviewee

Uh-huh. Can you expound when you say software architecture design techniques?

00:17:40 Interviewer

It can be microservices client based client.

00:17:44 Interviewer

Server or?

00:17:46 Interviewer

Something like.

00:17:47 Interviewee

So sure. So with my current company, I mean it's usually a mix, especially with banks, it's usually a mix because there are legacy systems, there are monolithic approaches to the software engineering and what what what we're trying to do is is to kind of modernise.

00:18:06 Interviewee

Those legacy implementations. So it's kind of.

00:18:12 Interviewee

I would say it was still monolithic, but we tried we we tried to implement microservices approach in every in every way we can but realistically it's very hard to to translate existing monoliths.

00:18:31 Interviewee

To more microservices approach, yeah, with the banks I would say, yeah.

00:18:37 Interviewer

Yeah. OK. OK. Yeah. So, which common software architecture design techniques of machine learning you found been using most completely your experience?

00:18:48 Interviewee

So sorry, I'm not really much of an Academy, so when you say software architecture, design techniques.

00:18:56 Interviewee

Well can.

00:18:57 Interviewee

You give me some examples.

00:18:59 Interviewer

For example, you have as you have you work in different companies. So maybe in one machine in one company where you work in machine learning and they use different software texture design techniques. Maybe they use client server technique. Others are using microservices, others are using different kind of.

00:19:15 Interviewer

You know, so in that perspective like.

00:19:18 Interviewer

Through your through.

00:19:18 Interviewer

Your experience with previous companies that you have worked how like which common software texture techniques you found that OK, these are the same like you you now for example you join the new company now. Now you found that OK, the same techniques that are being used so.

00:19:35 Interviewer

You have common analysis that.

00:19:38 Interviewee

Uh, yeah, I I would tell you, I would tell you that most in the industry.

00:19:48 Interviewee

How do I let's categorise?

00:19:50 Interviewee

So if you talk about, let's say startups or tech companies, microservices is really the most common way of how people build things.

00:20:01 Interviewee

Makes you know everything reusable, everything you can interact with with, with, with different kinds of features of of the programme of the application in in, in various ways, but those are typically being done in, in, in small companies. So for for for huge companies it's still very traditional.

00:20:21 Interviewee

And and that's the part where there's a lot of influx of dot engineers, machine learning engineers in in these huge companies because.

00:20:30 Interviewee

We are trying to modernise all those legacy implementations. We're trying to make them more robust, make them, yeah, event driven.

00:20:38 Interviewee

So yeah, in in some context, you know, small companies, they're small and tech companies, they're very much modern in the way that they do things, big companies.

00:20:50 Interviewee

They're hiring a lot of engineers right now to to modernise their existing uh infrastructure and existing and existing systems. Yeah, yeah.

00:21:00 Interviewer

So towards the 11th question, I think according to your experience, what are the best software architecture design techniques for machine learning and what are the benefits of using them so?

00:21:13 Interviewee

I I'm I'm I'm.

00:21:14 Interviewee

A client guy, so I'll be telling you that from a business standpoint.

00:21:22 Interviewee

The most beneficial technique is the one that gives value to the business users.

00:21:29 Interviewee

Business doesn't care about how it's being implemented. They only care about the results. They only care about.

00:21:38 Interviewee

How how it improves their bottom line?

00:21:42 Interviewee

So that's always what's important, whether you're it.

00:21:48 Interviewee

Team is agile enough, well versed enough to be able to satisfy the business requirements using a complex microservices approach versus a traditional software engineering approach.

00:22:05 Interviewee

I mean, whatever gives you know.

00:22:09 Interviewee

To the business, it doesn't matter so long as you give me the results. You know, I'm OK with that.

00:22:14 Interviewee

And and from from for huge for huge companies, that's normally how it works. But again tech companies, small startups, they're very they're very specific to technical approach. So yeah.

00:22:27 Interviewer

Yeah, yeah, exactly. So for the business perspective, the organisations try to focus like the results and the.

00:22:35 Interviewer

The system is workable, yeah, but but I think now I'm trying to emphasise for, for developing for the machine learning system.

00:22:45 Interviewer

You know that many developers and architectures and academic people now like which are the best common software architecture technique?

00:22:51 Interviewer

For machine learning.

00:22:52 Interviewer

Are being used to get the workable system.

00:22:56 Interviewer

So yeah, for example, if we say the normal software development life cycle, we say the class diagrams helps a lot for developing the for getting the requirement and and develop the accurate system.

00:23:10 Interviewer

So yeah, that was my my.

00:23:15 Interviewer

My point to emphasise with this point.

00:23:18 Interviewee

Uh, I'm not so specific about the the like the various different techniques, but let me tell you that what usually works with the business is that you create a small POC of.

00:23:32 Interviewee

What you're developing?

00:23:33 Interviewee

It doesn't. It doesn't have. It doesn't have to be a full-fledged.

00:23:38 Interviewee

You know very well designed programme from the get go. So typically you create an MVP. Let's say you give a uh, you take one or two sprints to to create a prototype to create an MVP of your machine.

00:23:52 Interviewee

Running system show the results. You know how they feed the data, what comes back as a result, create that very small prototype in in 1/2 or maybe even three sprints. Deliver that and once the business is happy with it.

00:24:09 Interviewee

Then from there you can you.

00:24:10 Interviewee

Know you could go.

00:24:11 Interviewee

Full-fledged you know, create your class diagrams, create your proper software design.

00:24:17 Interviewee

Design all your different classes you know. Let's say we we used a lot of Java in my previous company.

00:24:25 Interviewee

You know, making use of different APIs across the different business units. So you know you could you could you could go full-fledged on that so?

00:24:37 Interviewee

I think no. No one's stopping you from from being able to to go really technical on how you implement it, but the most important thing is you create that MVP. You create that prototype beforehand.

00:24:49 Interviewer

Yeah. Yeah, that's clear. So do you have any recommendation for software architecture design techniques for machine learning system?

00:24:58 Interviewee

Again, I'm not so specific about software architecture design, but from a from a machine learning use case standpoint, I think the most the most the.

00:25:11 Interviewee

The biggest the biggest. UM.

00:25:16 Interviewee

Component right now for you guys surprises.

00:25:19 Interviewee

Labelling the data.

00:25:21 Interviewer

OK. Yeah.

00:25:22 Interviewee

So, so, uh, labelling the data and having that fast end to end cycle of.

00:25:31 Interviewee

This is the data, you know what's the result? Optimise the model that that, that, that cycle, that that fast cycle is the is the one of the things that are that are missing because machine learning projects in in huge companies it's taking maybe between three to six months.

00:25:52 Interviewee

Just to get through that.

00:25:53 Interviewer

Yeah. So towards the 13 question, what would be the best practise that could be helpful or useful in applying machine learning architecture designing?

00:26:01 Interviewer

Of machine learning.

00:26:05 Interviewee

Best practises could be useful in in applying software architecture design.

00:26:10 Interviewee

What should be the best file? Can be useful, helpful in applying software architecture. Desire will be the best practise.

00:26:25 Interviewer

It can be like it also depends upon the system is real time system is not real time. So we have to look that way or we can also say that.

00:26:40 Interviewer

Which waterfall models have to be used and right solutions and practical experience is better?

00:26:47 Interviewee

I'm trying to find a good context in the industry of.

00:26:51 Interviewee

How to answer this question?

00:26:58 Interviewer

Right architecture.

00:27:08 Interviewee

So let me focus.

00:27:11 Interviewee

On the engineering engineering part of of machine learning.

00:27:13 Interviewer

Yeah, sure.

00:27:17 Interviewee

Yeah, let me try this so.

00:27:22 Interviewee

We did so in my previous company. We we did customise.

00:27:27 Interviewee

And develop an existing framework for doing dot engineering and machine learning engineering.

00:27:37 Interviewee

I mean building the framework.

00:27:40 Interviewee

It is very much tied up to the.

00:27:44 Interviewee

To the architecture design that the architects have initially drafted.

00:27:51 Interviewee

So I mean, I would say, you know, uh, you know, when the daughter comes in, we would have, let's say, three different data layers within it.

00:28:01 Interviewee

We got the validated zone integration zone, access layer zone. You know you draught all these kind of data models, information architecture.

00:28:12 Interviewee

So you have all this design. You have all this. You have all these data models and.

00:28:17 Interviewee

Then the engineering framework has been built specifically to that information design to that to that architecture design now.

00:28:30 Interviewee

The huge problem is.

00:28:33 Interviewee

As you go through the life cycle or as as the company evolves from from on a year to year basis as it evolves from from 1 requirement to another as new people come in you know replacing other people blah blah typically these solution.

00:28:52 Interviewee

Designs these solution frameworks.

00:28:57 Interviewee

Are getting updated or sometimes getting replaced so.

00:29:02 Interviewee

The problem was we have this existing framework. You know it's a tooling, it's a, it's an engineering framework, it's a dot engineering framework which is very much tied.

00:29:12 Interviewee

Up to this.

00:29:14 Interviewee

Information architecture that was built. The problem is this information architecture. Did this dot architecture.

00:29:22 Interviewee

Has since then been enhanced or has since then been replaced, so to speak.

00:29:29 Interviewee

Which rent, which is which, would which will render this existing framework like not usable at all? I mean you would need to do a lot of changes to this framework just to just to accommodate it, just to, you know, get through all those changes that has been done on the information on, on the, on the DOT architecture. So what I would like to say.

00:29:49 Interviewee

Yeah, when you build engineering frameworks, when you build machine learning frameworks, sometimes you know you get overly technical that oh, it has to be very specific to this, to this architecture blah blah.

00:30:03 Interviewee

But actually, your perspective in building that framework has to be, you know, in a very.

00:30:12 Interviewee

The the the customer that you should have in mind is actually the developers, so you don't necessarily have to design it.

00:30:19 Interviewee

You design the framework specific to the information or the architecture. You have to design it for how the developers will use it. So taking into consideration, for example, if the architecture changes if all the different technical stack changes.

00:30:32 Interviewee

Is this framework still going to live?

00:30:35 Interviewee

Because you've customised it so specific to that to that architecture. If something changes in the architecture, if one technology suddenly changes ohh we're not using Hadoop anymore.

00:30:46 Interviewee

You know, we wanna use Snowflake or we wanna use something else that this framework will suddenly become useless. So the the.

00:30:55 Interviewee

The the thing about designing the software software engineering principles is that whilst it's very useful to be very specific on on how.

00:31:05 Interviewee

On the design.

00:31:08 Interviewee

I think the most still the most important thing is versatility of the developers being able to.

00:31:15 Interviewee

Create something like like design something and you know be versatile. Be agile depending on on what the architecture will be.

00:31:25 Interviewee

Otherwise, your framework you know it, it will only be there, maybe for a good maybe two or three years after that. You know, all this stuff has changed.

00:31:35 Interviewee

You got to create another framework, yeah.

00:31:36 Interviewer

Exactly requires more time, yeah.

00:31:39 Interviewer

So towards the 14 question, what are the most common software architecture design challenges in machine learning? I think you have tried to explain something labelling data.

00:31:46 Interviewee

Yeah. Yeah, it's. I kind of touched that touched on.

00:31:49 Interviewee

That too, yeah.

00:31:50 Interviewer

Yeah. So if it remains the same, yeah. So what are the, I think this one is one of one of the important question of ours because we haven't found such questions in in the literature. So what are the main architecture decision on software architecture, design of different machine learning systems?

00:32:10 Interviewer

Like the major architectural decisions that you take.

00:32:13 Interviewee

Interesting. Interesting so.

00:32:17 Interviewee

Again, I would go high level first, OK. The the perspective I mean before designing anything the perspective, the, the, the perspective always starts with people, process and technology.

00:32:19 Interviewer

Yeah, sure.

00:32:29 Interviewee

So #1.

00:32:32 Interviewee

People, what kind of skill set does your people have? You know, do you have a people full of UM?

00:32:40 Interviewee

******** Java programmers do. Do you have? Do you have people you know who are mostly Python developers? Do you have people, mostly SQL developers?

00:32:50 Interviewee

So like, you know, specific to a company, you always have to cater it to what the skill sets of your people are.

00:33:01 Interviewee

Again, maybe if you're talking about, let's say, from an academic perspective, you know, oh, always, you know, the most technical, the most elegant solution.

00:33:13 Interviewee

Like I don't even need to care about people. I just need to care about process and technology and, you know, whatever is the best, whatever is the most elegant, you know, that's that's the only.

00:33:21 Interviewee

One that we consider but.

00:33:22 Interviewee

In the industry, you always have to consider people as number.

00:33:25 Interviewee

1 #2 process. You gotta understand what kind of.

00:33:32 Interviewee

What kind of uh processes that your your current company has processes in terms of what kind of development methodology are the guys doing? Are you in a completely agile environment?

00:33:47 Interviewee

What's the change management like if let's say you know you want to change code, is your cicd, you know very robust like you can change code deploy code as frequent as possible or it goes like three or five days before approval gets out blah blah. So processes is also very important.

00:34:07 Interviewee

And then lastly is about technology. So technology, they're different.

00:34:11 Interviewee

Years to it. So in terms of technology, number one is the.

00:34:18 Interviewee

Business. Sorry, I I think I forgot one thing. What one more thing. Business requirements. We also have to consider it.

00:34:26 Interviewee

There I would say that the most common problem for technical people is that.

00:34:32 Interviewee

If the business is.

00:34:33 Interviewee

Apartment is, you know, you're you're in the timeline. So let's say, oh, you gotta do something within three months, so obviously you won't create a very robust framework.

00:34:44

Or or or or.

00:34:45 Interviewee

Or architecture within three months.

00:34:46 Interviewee

You just need to get the data to the business blah blah. But yeah, so after the business is the technical stack. So on the technology.

00:34:54 Interviewee

Right #1 is what kind of technical stack do you have?

00:34:59 Interviewee

There's no correct answer of what of of what's the single most? What's what's the proper? Uh, well, what's one?

00:35:07 Interviewee

There's no one tool that, that, that that is recommended for for each and every company. So it always it always depends on the situation and there's a lot of various factors around that, but very important to consider what technical stack.

00:35:18 Interviewee

Do you have? Do you have? Do you have it on cloud? Are you still on Prem? Blah blah.

00:35:26 Interviewee

#2.

00:35:27 Interviewee

Is sorry, what are the main architectural decisions?

00:35:32 Interviewee

#2 is your information architecture. It's very important that you design your machine learning system. Depending on what is, what is the existing information architecture, where are you gonna get the data? Is it already cleansed data well?

00:35:51 Interviewee

In in which kind of format are you going to receive it so you know all about the the information in data?

00:35:58 Interviewee

Sure. Another thing about the decisioning is what kind of front end systems are you gonna have. So again, depending depending on the machine learning application that you're building, let's say some of them, you're going to deploy them into some websites or or some campaign campaign applications. So it's also very important to understand what kind of.

00:36:20 Interviewee

Well, what what are different channels? What what are different kind of front end?

00:36:23 Interviewee

Applications that uh, that, that.

00:36:26 Interviewee

Your machine learning application is is going to communicate with.

00:36:30 Interviewee

So once all of that landscape has been set.

00:36:34 Interviewee

Then you can go and decide as an architect software architect. Then you can go and kind of draught and decide, OK, how am I going to build this application? You know? Am I going to design A for example, now I'm going to give you an example.

00:36:53 Interviewee

Am I gonna am am I?

00:36:54 Interviewee

Gonna design A completely event.

00:36:59 Interviewee

Event based or microservices driven kind of machine learning application. When all the systems that I'm communicating with are legacy systems you know are very monolithic systems and you're trying.

00:37:10 Interviewee

To build a.

00:37:14 Interviewee

You know, a very modern kind of application, but you're trying to communicate it with with monolithic.

00:37:20 Interviewee

You're not going to do that, right? So I mean, I kind of went through a lot of things.

00:37:26 Interviewee

Being able to to, to, to kind of print out the landscape as an architect, then you would be able to see oh.

00:37:33 Interviewee

It's right there, you know, I now I know what kind of software design that I can build across all of these. Again, maybe this is in the context of.

00:37:45 Interviewee

Trying to implement some some modern stuff, some machine learning.

00:37:48 Interviewee

Stuff into an organisation that is so huge.

00:37:52 Interviewee

The It's it's a totally different context when you're a startup, you know you're trying to build a product entirely.

00:37:57 Interviewee

New brand new.

00:37:59 Interviewee

Let's say there's a block. There's a lot of blockchain happening right now. Yeah, you try. You wanna, you know, add some some machine learning apps into it.

00:38:04 Interviewee

Then of course you can go microservices. You can go very, you know, modern based approach APIs and all.

00:38:12 Interviewee

Like like you can do that. So it it it's it's just that when the context becomes really different when you when you're in that kind of environment versus when you're in a in a in a, in a, in a in a in a company such as banks or telcos or something.

00:38:38 Interviewer

Yeah, yeah, I think I have learned a lot from your experience.

00:38:46 Interviewer

And your expertise, it was really a very beneficial session for me.

00:38:51 Interviewer

Yeah, yeah, it will. Definitely. Yeah, it will definitely help. Yeah, I think I have. You have like some very interesting topics and.

00:38:51 Interviewee

Do you think it helped?

00:39:02 Interviewer

Yeah, yeah.

00:39:05 Interviewer

I don't know how to explain what the I have learned, so yeah.

00:39:11 Interviewee

Thank you. Thank you. I mean you, you mentioned that you were currently you're working uh in in C#. So are you are you trying to make that jump into machine learning and AI?

00:39:23 Interviewer

Actually, I have actually when I have done my bachelors I have work in machine learning. I have developed like finding the age, gender and mood detections.

00:39:38 Interviewer

Through the real time analysis. Yeah, and.

00:39:40 Interviewer

You some other like some other small projects like text detection or something. So that was my interest in that after bachelors I have bought some projects not very highly and but then I started to work in the website development.

00:40:01 Interviewer

So yeah, I have a plan in the future to work in the machine learning, but let's see how it goes.

00:40:10 Interviewee

Yeah, yeah, I mean.

00:40:11 Interviewee

The machine learning space is kind of.

00:40:14 Interviewee

I would say kind of interesting these days because.

00:40:20 Interviewee

I think about two years ago, you know, there's a two, maybe two to five years ago, there's a lot of interest in machine learning, although I think the past one or two years the focus has mostly been in data engineering.

00:40:33 Interviewer

Well, yeah, that is fine. We can see.

00:40:36 Interviewee

No data science was there from the start, but the past couple of years the interest the focus has been a lot on dot engineering, and this is mostly because of companies trying to modernise their data platforms.

00:40:43 Interviewer

OK.

00:40:48 Interviewee

Hmm, after maybe I would think that after the next two to three years, it will go back to machine learning, but I'll be very interested to see.

00:40:56 Interviewee

In which state the machine learning is? Because have you heard about auto, ML?

00:41:03 Interviewee

Auto machine learning.

00:41:05 Interviewer

Yeah, auto machine, yes.

00:41:07 Interviewee

Yeah, yeah, auto ML.

00:41:08 Interviewee

So there's a lot of vendors who's was is kind of recommending this auto ML tools. We're in it kind of removes those specialists like ******** data scientists, ******** machine learning, machine learning people.

00:41:27 Interviewee

Uh, because these tools makes it easier for you to to use any models that you want, you know, you just feed the data, you know. Uh.

00:41:37 Interviewee

Just use the tool.

00:41:38 Interviewee

And it will automatically use the best algorithm and you know the model is right there in front of you. Yeah. So it's very interesting to see how this machine learning landscape will turn out to be.

00:41:52 Interviewer

Yeah, that would be interesting. Yeah. Yeah, yeah. Also work in the BI system-based intelligence system, the company. So yeah, the companies are very interesting nowadays to play with the data and try to extract some useful information from them.

00:42:07 Interviewee

Yep, yep.

00:42:09 Interviewer

Yeah. So, I will share with you this recording soon enough. I will try to upload and then we'll share you the Google Drive link so you can have it.