Transcript

00:00:00 Interviewer

Yeah. So I'm starting to recording this session. After getting the permission from you so we can start by from start from the question one.

00:00:11 Interviewer

Can you please introduce yourself and describe your job role in this company?

00:00:16 Interviewee

Yes, OK. My name is " ".

00:00:21 Interviewee

And I'm from " " and my company is " ".

00:00:28 Interviewee

Uh, and my role is the main uh developer machine learning engineer.

00:00:39 Interviewee

And the managing the scientific context for for the IT is, it is a startup startup. So as you know, in startups one man is doing team jobs.

00:00:59 Interviewer

OK. Yeah. Nice. So since how many years you are working in this company?

00:01:04 Interviewee

Uh, since the start of this year.

00:01:08 Interviewer

Start of this year, OK, so how many total experience do you have?

00:01:15 Interviewee

Or this company or in general?

00:01:18 Interviewer

In general.

00:01:19 Interviewee

Ohh Angela, I before being uh or transfer a transaction to a machine learning engineer, I was a a software engineer. I worked with large companies like Upland Software in US.

00:01:36 Interviewee

With large projects up there like power steering and this application is dealing with the broad and the large customers in in the world and starting from.

00:01:56 Interviewee

Machine learning engineer I'm I'm working.

00:01:59 Interviewee

And and several projects for several companies like aggregate Genius Mines in in Canada.

00:02:11 Interviewee

But in in this company, my focus or my domain is in bioinformatics chemoinformatics work.

00:02:21 Interviewer

OK.

00:02:22 Interviewer

Yeah. So total, how many years of experience do you have becomes?

00:02:27 Interviewee

General total will be like 11 years.

00:02:30 Interviewer

OK, 11 years. Thanks. So.

00:02:35 Interviewer

OK, so towards the third question, have you any published any thesis in machine learning?

00:02:41 Interviewee

In machine learning, uh, not yet. But I'm tending to maybe next year.

00:02:46 Interviewer

OK.

00:02:50 Interviewer

So can you please share your experience in in the current job?

00:02:55 Interviewee

Uh, yes, currently my main that male language I'm using is R.

00:03:05 Interviewer

OK.

00:03:06 Interviewee

So I'm trying to reach the the language with many packages as I I could, as I said my business domain is chemo formatics and bioinformatics so I'm trying to create.

00:03:27 Interviewee

Life cycle packages for it assigns for these two kind of business in art.

00:03:35 Interviewee

Starting from gathering data, cleaning data until I can build the models and verify this that model.

00:03:50 Interviewee

I have published packages already from since two years and.

00:03:57 Interviewee

I'm going to publish more in few in next few months.

00:04:02 Interviewer

OK, nice.

00:04:04 Interviewer

So do you have any experience in the previous company which is developing machine learning system? If so, what was your only space? I think you have already.

00:04:11 Interviewer

Said you don't have.

00:04:13 Interviewee

Just resting.

00:04:16 Interviewee

Not not developer, but I started developing.

00:04:22 Interviewee

In my company.

00:04:23 Interviewee

Nothing. Put my briefcase for.

00:04:25 Interviewer

OK.

00:04:27 Interviewer

Yeah, towards the 6th question is your company service based or product based?

00:04:35 Interviewee

I'm not providing services, but I'm building packages so I'll say it's product based.

00:04:42 Interviewer

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

00:04:49 Interviewee

I'll try it.

00:04:52 Interviewer

So could you please share experience with the interesting projects in machine learning that you have worked on recently?

00:05:00 Interviewee

Uh, yes. Uh, I have a big.

00:05:05 Interviewee

Project called the drug verse.

00:05:09 Interviewee

Drug verse.

00:05:09 Interviewer

Drug wears OK.

00:05:12 Interviewee

This project actually is a collection of a project. It's like it is.

00:05:21 Interviewee

It it tries to mimic Tidyverse in arch.

00:05:25 Interviewee

Uh, but for drugs? Uh, so it contains multiple packages inside it. Inside it there are three published packages. Until now person for.

00:05:43 Interviewee

Extracting drugs environment data from public databases like drugbank and convert them into multiple data frames or data sets that the researcher can work on on them.

00:05:59 Interviewee

Also, there is a coffin 19 band and it is a database data set that contains information about candid drugs.

00:06:13 Interviewee

That can be used or or it is used right now by researcher for finding a cure for COVID-19.

00:06:22 Interviewer

OK.

00:06:23 Interviewee

Uh, and there are similar package for like that but for uh, a wide a wide range.

00:06:34 Interviewee

Of uses for drugs.

00:06:36 Interviewee

And currently I'm containing on three packages under the same project, one called DVDV Hotel DV for drivers and hotels provide utilities for.

00:06:56 Interviewee

For drugs development researchers to work on drugs data in particular, like ranging the data build extracting features from drugs or from targets building.

00:07:17 Interviewee

Characteristics, matrices, etcetera, from from drugs development perspective and the DV blot which is is specialised in blotting drugs and routines information.

00:07:36 Interviewee

All of that in and using R and for plotting. I'm using a library called Campus Express.

00:07:45 Interviewee

For for plotting.

00:07:47 Interviewee

And the third package is called DV Bread and it is for elementis multiple machine learning algorithm dedicated.

00:08:00 Interviewee

For drug development.

00:08:06 Interviewer

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

00:08:12 Interviewer

You work with.

00:08:15 Interviewee

OK as my my domain is not not usual and it's not normal, it's it's business not not normal business like any other business in in.

00:08:32 Interviewee

Machine learning. So there's actually no one rule can be applied on that. Even even the algorithm cannot be classified easily. So I'm we are starting building our process there.

00:08:48 Interviewer

So which common software architecture, design techniques of machine learning you found being used in most country through your experience?

00:08:56 Interviewee

Uh, trees, actually, we are mainly trees. Great.

00:09:02 Interviewer

Trees gate.

00:09:04

Right.

00:09:05 Interviewer

Sorry. Trees, trees, gate.

00:09:07 Interviewee

Grid searching we we are building grids of many models and trying to find the best of them.

00:09:15 Interviewee

And and also we are trying to use natural language processing techniques but we are trying to rub them to work on our data because current tools are working on natural language processing for humans language.

00:09:36 Interviewee

Which is totally different than we have the data we have.

00:09:42 Interviewer

So are you using any architecture like microservices, client based services or anything?

00:09:48 Interviewee

Like what? Sorry.

00:09:49 Interviewer

Are you using any architecture like like client, microservices, client, client, server services or something like?

00:09:58 Interviewee

I'm using sometimes I'm using API's or services from other bio developer.

00:10:10 Interviewee

But not big fancy thing, because it's the whole thing. For, for, for, for the pub perspective is a little new and and a little not very popular.

00:10:23 Interviewer

Yeah, I can understand. Yeah. So according to experience, what are the best software architecture design techniques for machine learning? And what are benefits?

00:10:32 Interviewer

Of using them.

00:10:34 Interviewee

Uh, for for for drugs? Uh, development. In particular, the best technique we are using great searching because it's it allows us to build multiple models and have we are having a lot of.

00:10:54 Interviewee

Parameters for each algorithm and the small initialization for each prompter can.

00:11:04 Interviewee

Make us have a very different result, so having a great.

00:11:13 Interviewee

Can allow us to have the best model.

00:11:18 Interviewee

And and also help us in verification process.

00:11:24 Interviewer

OK. Yeah. So do you have any recommendations for software architecture, design techniques of machine learning system?

00:11:34 Interviewee

Or for tools or techniques or or what?

00:11:37 Interviewer

Software architecture design techniques. Any recommendations for that?

00:11:43 Interviewer

Like when you design the software architecture or like do I need design the machine learning systems? Do you have any recommendations for that?

00:11:55 Interviewee

Not, not, not really, not really, actually, just the standard standard process.

00:12:02 Interviewer

So what would be the best practise that could be useful or helpful in applying software architecture, designing of machine learning system?

00:12:14 Interviewee

Or that will be the first step cleaning your data. Have a clean data.

00:12:19 Interviewer

OK. Yeah. Yeah, OK. You need the data.

00:12:23 Interviewee

And also make sure that data is not not only clean but also complete in a way that the algorithm will not will not break, will not be biassed.

00:12:41 Interviewee

So you need you need your data have to be diverse. I have diversity information as much.

00:12:47 Interviewee

As you could.

00:12:52 Interviewee

All all textual data, categorical data are included correctly in manner you can.

00:12:59 Interviewee

Make use of them.

00:13:04 Interviewee

Any any any recommendation will be and and the data wrangling and feature engineering steps they are the most important.

00:13:14 Interviewer

So what are the most common software architecture design challenges in machine learning systems?

00:13:23 Interviewee

OK.

00:13:25 Interviewee

And in my in my state.

00:13:28 Interviewee

It was having not too much memory. We have a very huge data.

00:13:35 Interviewee

To work on.

00:13:35 Interviewer

OK.

00:13:37 Interviewee

Like a.

00:13:41 Interviewee

Tiers of data and.

00:13:46 Interviewee

And it was a challenging not only actually in in in, in, in, in machine learning you you might use well but working on in on.

00:13:56 Interviewee

Running the models is is most easy, easiest part, but the the most difficult one is to build.

00:14:05 Interviewee

Visualisation techniques to extract feature engineering from from data.

00:14:10 Interviewee

From this data.

00:14:13 Interviewee

Especially that chemo chromatics and promatics data are not common and most of current techniques.

00:14:24 Interviewee

Cannot be applied directly on them. We have. We had to the to do a lot of transformations.

00:14:33 Interviewee

And the cleaning to make them usable and.

00:14:38 Interviewee

Also build our algorithm to visualise and extract useful information from this data. That's why we are trying to build.

00:14:47 Interviewee

This track versus project.

00:14:51 Interviewer

OK. Yeah. So what are the main architecture decisions on software architecture, design of different machine learning systems?

00:15:00 Interviewer

Like the major are that your decisions that take?

00:15:05 Interviewee

The major decision for us were to make our.

00:15:09 Interviewer

Yeah, yeah.

00:15:11 Interviewee

Or to make our own packages to mimic the data science lifecycle.

00:15:20 Interviewee

We will not depend on the usual labels.

00:15:31 Interviewer

OK, so your it it will be like if you have a new system or maybe if you don't have find some libraries it will be nice if you make your own packages and make your own libraries.

00:15:47 Interviewee

Yes, if you are your what you are doing.

00:15:52 Interviewee

Is repetitive or someone is will do the same thing. Make your work on a package or a.

00:16:00 Interviewee

Library and share it.

00:16:02 Interviewee

That's that's.

00:16:04 Interviewee

Yeah, I can tell.

00:16:06 Interviewee

You one of the libraries that is a life example for that.

00:16:13 Interviewee

If you want to use it in your dirt.

00:16:17 Interviewer

Yeah, sure.

00:16:20 Interviewer

Yeah, you can handle it.

00:16:22 Interviewee

I'll I'll post it on.

00:16:24 Interviewee

The chat.

00:16:25 Interviewer

Yeah, sure.

00:16:27 Interviewee

This is the.

00:16:27 Interviewee

First package from drug verse.

00:16:34 Interviewee

You can you. Yeah, you. It's have a statistic. It's on a statistics et cetera you can.

00:16:41 Interviewer

OK.

00:16:43 Interviewer

Yeah, I I will have a look there.

00:16:49 Interviewer

Yeah. So.

00:16:52 Interviewer

We have 15 questions, so I think we have answered 15 of them.

00:16:59 Interviewee

I hope, I hope I could make any.

00:17:02 Interviewee

News for you.

00:17:03 Interviewer

Yeah, your your answers will be very beneficial for me when I try, when I will try to compile all the answers that I have collected through the interviews.

00:17:13 Interviewee

Great. OK.

00:17:14

Thank you.

00:17:16 Interviewer

Yeah. So if you have any questions, then you can ask and.

00:17:23 Interviewee

No, no.

00:17:24 Interviewer

OK. Yeah. So thank you so much for your valuable time. I really appreciate it. So welcome.