Interview 37 - Amin

Interviewer 1

All right. Um, so to start off, can you give us a bit of information about yourself and how much experience you have in general and specific to machine learning?

Interviewee

Yeah. Okay. My name is Interviewee. Uh, I'm working. Uh, well be, right now I work for myself, but, uh, some, a couple years ago I used to work for a company here in Country X, which they, they make large engineering projects for the infrastructure market, mainly high.

Interviewee

Uh, High voltage, uh, electric machinery. And right there I had the opportunity to work with machine learning firm. Well to, I'm an engineer, I'm mechanical engineer for my primary formation. And, uh, I used to to make lots of calculations about some machine, uh, performance mainly. For security, for efficiency, for, for health or, uh, even, uh, environment stuff.

Interviewee

And because of that, most part of this, of this calculations and as we, we used to, to work with engineered engineering system all the time. It's, it's, I had, we had some different per parameters and you know, this used. Have a lot of ifs and, uh, for in inside our code. And because of that, uh, of course, It's, uh, it's became too hard to manage this, this, the system.

Interviewee

Uh, cause like the quantity of Fs, uh, sometimes used to be, my God, 50 70. Impossible to put your hand on, uh, a system like that. And because of that, I used to, I used to, I started to, to. Machine learn for, mainly for, uh, regression problems to the regress. We had lots of features, uh, for the point of view of design and usually we, we need to reach some, uh, what was our goal was to get some.

Interviewee

Wow. How, what, what does, how does machine has the, the higher efficiency, for example, how, what does machine has low points of copper used for construct, uh, things like that? Um, because of the start to, to chop apply, uh, the first was of course, uh, linear systems, uh, using linear system and. And sometimes this, um, well linear system, you know, , we don't, we, you don't have a, you usually don't have linear system inside of our, our seldom, not never, but seldom we have, uh, linear system and, um, and we start to, to.

Interviewee

Gradually to no linear. And today I am most part of my time. I am, I'm training, uh, neuro network using 10 intensive flow, uh, carers more today than intensive flow. Um, that's the, that's the, the, the whole idea.

Interviewer 1

Yeah. Nice. So it's great to have you.

Interviewee

Yeah. Um, yeah, go for it. No. No. Yes. Uh, and beside that, I, I, I, I, I, I have some, I'm curious, guy , and, uh, I have some, some stuff in, in the financial market as well.

Interviewee

I write some, some trading systems, mainly for Country X assets. And, uh, that's the, the. You today? I'm, I, I'm, I spend much more time just in this financial stuff than engineering stuff. .

Interviewer 1

And you build model for, for the finance, uh, uh, you build model for,

Interviewee

okay. Yeah, yeah, yeah. I had, uh, we have system that we have a system that gets information from market, uh, some Forex or an OR, or, uh, stocks.

Interviewee

And you read this information and from that, we had, we had, we tried. We try to get some well patterns, try to, oh, this time is, this is time for, for buy, this time for, for sell, this kind of thing. Interesting.

Interviewer 1

Uh, so to start off, I will ask you a general and open-ended question. So what are the main quality issues you have encountered with your data model or system so far?

Interviewee

By far, by far is data by far. Because today, today you have the second learn from Python and. You can, you can, you can tune your parameters and get a really good, a really good, uh, uh, uh, precision. And, uh, but the, the by far is the, is the data because I dunno how it works in, in Country Y, but our. I'm the financial marketer here in Country X, the financial market here in Country X.

Interviewee

Uh, for the most part of the time, this is really regulated, but for some stuff they're not. And, uh, this guy, the brokers here, they, they, they, they, they understand the, the, the, the easiest way they, for the, they get value. They, they get money. To provide liquidity for their customers. This liquidity does not came from the market, came from themselves.

Interviewee

So this a kind of noise. and for some, uh, and depends on the, the kind of broker you're using. If they works really with the, the, the main part of their, of their customers are, uh, them money or personal individuals. They inject a lot of noise inside this, this, this data. And they, they, they provide for every.

Interviewee

And because of that, uh, we, we need some kind of other source of, of market information here. And, uh, it's difficult because, uh, this other source are not, uh, friendly for the system, the most part of the system we have. And because of that, we need some, some engineering, some data engineering, and. And this, it's time consuming, and the reli, the reliability of this of it is difficult to, to, to check as well, because as you know, just a bunch, a bunch of data.

Interviewee

Uh, every day is a bunch of data and you, you need some, some way you need to confront the boss to see . Uh, if you, if you are getting. How is the quality of the, what is the difference between them and how is the quality, uh, one compared to each other? And this, this is, this is taking a lot of time of us and because, and, and.

Interviewee

It's for, for nation market, for the engineering point of view, the problem is the same because we have sensors inside of our machine, the sensors cap, uh, they, they cut the information, uh, they get information necessary for, to us to, to make our, our, to tune our models and this kind of thing. Well, , uh, sometimes this is dis calibrated.

Interviewee

Sometimes you, some, uh, if, uh, someone forgot to, uh, something with, if, with, uh, uh, With some heating source close to our perimeter, uh, gathering, uh, noise from, uh, not expected source. Well, a lot of, a lot of, all sorts of noise, uh, you get from, from, from this side. And again, we need to filter. This data to, in compare to what you think is a, this is a good quality data and we need to compare this, those two things to, to not, to have a, a idea of how good or how bad is our intake data is coming in.

Interviewee

I

Interviewer 1

see. Thank you. I wanted to ask you about the data for financial system. You mentioned, uh, liquid liquidity. Basically, I, I do not have a big background in finance, so I just Okay. Can you explain me what is the issue with the data? Um, someone doesn't

Interviewee

know, uh, okay. Finance that much. Yeah. Well, liquidity.

Interviewee

Liquidity means, for example, for you is when, right now, if you want, want. If you want to buy, for example, a stock from Apple, it's really easy. You can open your broken system There is a billion of people trying to, to sell for a specific price, uh, the Apple stock. Uh, for you, this is not true for. For every asset we have inside, uh, for any stock market in, uh, in the world.

Interviewee

And when you decrease the quantity of people want you buy and people want you to sell, usually you have higher prices. This is this, uh, kind of, uh, demand, this economic stuff, uh, and, uh, just, just lack of, just lack, just lack of quantity of buyers and sellers. It is what our, uh, in the market, this call liquidity, for example, uh, uh, you want to, for example, you want to buy, for example, you want to buy a Ferrari.

Interviewee

How many people in the world has Ferari to sell? Compare for any Volkswagen or Toyota or Honda? That's a good idea. What is the difference between liquidity Toyota and Honda? Volkswagen is high liquidity, uh, whereas , Shar Bug, uh, McLaren, this kind of stuff is really. Really, uh, less quantity available to, to trade.

Interviewee

I dunno if I ask. Was clear. Yes,

Interviewer 1

yes, it's clear. Thank you. Okay. Right. Uh, so moving on to data collection. I will, uh, well basically I will ask you about three data collection process and I just want to know if you ever use it and if yes, if you ever had issues with it. Uh, so did you ever have data that was manually collected?

Interviewer 1

So it could be, for example, someone who labeled your data or someone who really sampled the data? It it writes down the values. Yes.

| Data-collection-manual | Data-unnormalized-annotations | Data-unnormalized-annotations-fix |

Interviewee

I work with manually collected data. It's terrifying. . It's terrifying. It's horrible. It's horrible. Yeah. Well, ma mainly if you in the engineer at the engineer, it's much more easy.

Data-collection-manual   Data-unnormalized-annotations   Data-unnormalized-annotations-fix

Interviewee

You say to the, to the guy, collect the data. Yes. Every time you ha you, you see this value due to the amount to write, the hears down. It's not difficult for engineering, but first other stuff, like for example, how was the day? What? How was the temperature of day compared to yesterday? This kind of categorical things, and this is the problem.

Data-collection-manual   Data-unnormalized-annotations   Data-unnormalized-annotations-fix

Interviewee

Uh, for, for example, uh, here in Brazil right now is summer. Yesterday was like 38 cel degrees. Here was my god was terribly warm. And for me, for my wife, for my son, for every, for everyone. Here at home, uh, at, at night, we are talking with our, with our. He lives in a neighborhood. Right? Right. Okay. Five kilometers from, from here.

Data-collection-manual   Data-unnormalized-annotations   Data-unnormalized-annotations-fix

Interviewee

And I'm sure there, the temperature was exactly the same for him. Was not ho was. Yes. It was not warm. , uh, incredible. Incredible. How, how can I say, per person, personal can, can be this kind of thing. And when you have. Manual, manual, collective data. This, this problem's very common. It's very common

Data-collection-manual   Data-unnormalized-annotations   Data-unnormalized-annotations-fix

Interviewer 1

for sure.

Interviewer 1

Okay. So if I understand correctly, basically there's, people are not subjective in your, in your experience, when they say simple data Yeah. They, they

Data-unnormalized-annotations   Data-unnormalized-annotations-fix

Interviewee

are subjective. Yeah, they are. Yeah, yeah, yeah. Mainly, mainly when you have numerical, numerical, That is no problem. The problem when you have categorical, categorical is the main problem.

Data-unnormalized-annotations   Data-unnormalized-annotations-fix

Interviewee

Oh, okay. I see. Thank you. At my experience of my experience

Data-unnormalized-annotations   Data-unnormalized-annotations-fix

Interviewer 1

that, yeah. And how, how do you address this issue? Usually?

Data-unnormalized-annotations   Data-unnormalized-annotations-fix

Interviewee

Yes. Normally what, what, what you try to do with, before the guy collect the data, we write a such a, uh, head, a guideline to make the fear as clear as. Them to him, to her. How is the way you need to, or, or you have to categorize this, this stuff.

Data-unnormalized-annotations   Data-unnormalized-annotations-fix

Interviewee

That's the, the, the main, well, normally what we, we, we used to do is, uh, exactly this. .

Data-unnormalized-annotations   Data-unnormalized-annotations-fix

Interviewer 1

Okay. And just, just to be sure, when you mean categorized, are you meaning about, are you talking about the labels or the features

Data-unnormalized-annotations   Data-unnormalized-annotations-fix

Interviewee

or both? The, the features. The features, okay. Mainly the features. Yes,

Data-unnormalized-annotations   Data-unnormalized-annotations-fix

Interviewer 1

the features. And, and so what you mentioned is that basically you're trying to prevent the problem, but once, if you have the problem, , do you have a way to detect that there is problem and fix them or not

| Data-unnormalized-annotations | | Data-unnormalized-annotations-fix |

Interviewee

really?

| Data-unnormalized-annotations | | Data-unnormalized-annotations-fix |

Interviewee

Yes, we try because, uh, naturally our ma, like I said, I never, uh, I never faced the problem with the model inside, uh, in, in itself how the problem we have came from our data. Every, every time I was debugging in my, my machine learning system , I ended up at the same, the same problem, the quality of data. And because of that, along these years, uh, I've been taking really care how, how the way we collect the data, the way we instruct people to collect data and, and.

| Data-unnormalized-annotations | | Data-unnormalized-annotations-fix |

Interviewee

Um, yes, because of that, because of that, we, and because this, it's the most, by far, this mo this most time consum consuming stuff I had to do. Uh, I, I, I tend to, I tend to spend a lot of time writing, uh, guidelines to the people to collect this, the.

| Data-unnormalized-annotations | | Data-unnormalized-annotations-fix |

Interviewer 1

Okay. I see. Thank you. Interesting. Okay. Uh, have you ever used data that w uh, well, external data, so public dataset, uh, third party API or, uh, web scrape data?

Interviewee

Yes. Uh, just, uh, for example, for the market. For the market, what we have is consumer api. Api, and like I said to you, because the, the peculiar stuff here in Brazil, we need to. We can, we can't, we can't, um, we can't, um, this, this data. We cannot simply, we cannot simply close the eyes and imagine this, this is a good quality data.

| Data-collection-API | | Data-collection-manual |

Interviewee

Yes.

Interviewer 1

Thank you. That's right. You already mentioned it. Yeah. Uh, and finally, have you ever used data that was generated by another system?

Interviewee

At engineering problems? Yes. Yes. Okay. Because yeah. Yes, we, we, we have some, it's more, uh, it's better. Usually it's better because, you know, you have the documentation, the people write the, the, the system and what they, what of stuff coming into the system of them and what the, the data come out and usually, By much better quality than anyone of the other sources.

Data-collection-system

Interviewee

For sure.

Interviewer 1

For sure. Okay. See, thanks. Okay. Um, moving on to data preparation. Have you ever measured the quality of your data and or tried to

Interviewee

improve it? Yes, all the time. , that's the, our main problem for sure. Our main problem, like I said, uh, What, what I had to do sometime. What? Right now what I do, I, I, I spend a lot of time trying to get high quality data and we make, uh, we make some, um, statistical analysis for qu uh, just say what the distribution of the data.

Interviewee

And, uh, all the time we had new data. We compare to this dis to, to this distribution. That's the, um, this distribution and the usually expected range of values as also. Uh, for numerical or for categorical features. And, uh, by the time we see in, uh, uh, naturally we have some, some, um, uh, liability fidelity, uh, value to compare with, uh, when the, when this value are, are overdrawn.

Interviewee

We, we, well stop. Stop. We had problem. , but yes, for sure .

Interviewer 1

I see. Thanks. Um, is there any other data quality issue we missed that you consider relevant? The

Interviewee

data collection or what kind? Uh, any data issue. Yeah. Uh, well, yeah, missing data. Usually it's common. Um, wrongly assign the data. It's common as well, uh, for mainly for manual collecting for, uh, um, automatic is less problematic.

Interviewee

The automatic we have. The problem is with other. Other nature, they had other nature and that the, and um, you know, our main quality today is, I would say, will be, we have, uh, data w can rely on without too much to too, spend too much time to see the qualities. Good.

Interviewer 1

Okay. And you ju you mentioned, um, that the quality, the nature of the quality issues and systematic data generated by system, uh, is different from manually collected data. Yeah. Yes. I just want to ask you, what, what is the difference? What, what are the type of issues in the

Interviewee

data? The main issues when you have, for example, noise.

Interviewee

Uh, they, the most part of time I had problem with, uh, automatized the system. Uh, cause they, inside, inside of this, this, this data, we have a lot of noise. Sometimes some kind of noise is spectrum or problem. But sometimes we have a lot of, uh, a lot of noise and this lot of nice can deplete the quality of our, our model.

Interviewee

I had this right, I I dealing with it right now, right now with the, uh, uh, a robot and I'm, I just wrote the, I just wrote it right now. I'm collecting. His precision and this kind of stuff and, uh, was good. But right now this precision is, is coming down. And, uh, naturally my code doesn't, does not change. So the real, the problem, probably the problem probably is the quality of the data.

MLP-unstable-external-dependency-fix

Interviewee

They, they, he's collecting, it's collecting right now. And, uh, the I think, uh, they, it's a bit more, a bit more than right now than used.

MLP-unstable-external-dependency-fix

Interviewer 1

Okay. And do you have any tool to, to detect issues and data and to fix them?

Interviewee

Uh, right now? Yeah. Yes or not, not. I, I wrote, uh, I wrote A V B A A V B A routine Hearing Excel.

Interviewee

And I collect some, some batch of data and compare, like I said, I compare to distribution in range of values. And when this distribution is skewed a bit for, uh, to left or right. I, I, yes, I know this. The injection noise. Yes. Okay. Uh, what I say no, I say no because this is stuff I, I. It's was, was not, uh, like I say, um, any, anyone else, uh, has checked if it's right or not.

Interviewee

That's the reason. Uh, I say not . Yeah.

Interviewer 1

Yeah. Thank you. And moving on to model evaluation, how do you evaluate the quality of your. And as a reminder, quality is not only defined by ML performance, so accuracy, influence, score, uh, you name it. But there's also other aspects such as explainability, scalability, efficiency, uh,

Interviewee

yeah.

Interviewee

Well, like, like a, sorry.

Interviewee

S problem, quality, scalability? I don't think it's a problem for us, for me, because I, I use, I use tensor floor, carers by, by torch, any kind of stuff. This is part of their work frame. Uh, I would say. And. The, um, the problem, the problem I face, uh, with this machine learning model usually has to do again with the data and.

Interviewee

What, for example, uh, I was trying to, to solve a problem here and I used to, to have some reference to the cargo data sets written form. For another people. And what I have noted is for it's a big problem, a big, big issue right now. A big issue. Big, big. This guy, they used to, to make a, what's called you, you, you had to, I think you know that you have, they.

Interviewee

What is calling statistics arbitrary, waiting for categorical steps a lot of, a lot, and the problem with it is even you using this, this wrong stuff, you can have a good precision quality for your machine. even using this, the, the, this wrong stuff. And I think today, this, this is the mo, the, the bigger problem because you have tested and mature system for, for chi tuning your hyper parameters and this kind of stuff.

Interviewee

But this, this same does not check what the of this arbitrary thinks. Happens to the data set that come into this, to this model .

Interviewer 1

Okay. Uh, so, so what you're saying is that, um, the score was weighted, uh, arbitrary arbitrarily. Am I correct?

Interviewee

Uh, sorry, I had a problem with the, your video. I didn't hear what you say.

Interviewer 1

Okay, no worries. Uh, so, so the issue you mentioned was that, Just, maybe I misunderstood. Uh, the samples were weighted differently from one category to another when they were computing the, the final score is, is that right? No.

Interviewee

No, not exactly.

Interviewee

Uh, for example, you have a, uh, liquid, a scale for something. Uh, I, I agree, disagree. This kind of stuck the kinda. What these guys do? Yes. Uh, today, usually. Oh, I agree. I they assign number one. I disagree. They assign minus one. I agree. I agree highly. They assign three and they simp put this numbers together and inject this inside the, the, the model simples like that.

Interviewee

Simple like that, that, and like I said, when you there, there's way to transform categorical features into, uh, into numbers. Mm-hmm. , um, for, we have a lot of technic technic stuff to technical techniques, techniques you can do in statistics to, to, to do that. But these people does not. Yeah. Yeah.

Interviewer 1

Okay. I see.

Interviewee

Thank you. Okay.

Interviewer 1

Have you ever assessed a quality of your model with the user of if it, if it's applicable to you, have you ever use, sorry, I will repeat the question. Okay. Have you ever assessed a quality of your model with the user of the system?

Interviewee

Yes. Yeah. Yeah. All the time. That's the main reason we, we construct something to, to deliver to somebody or for me or for anyone and, and to, to get value from this.

Interviewee

Uh, that's the, the reason, uh, that's the, the, that's the stuff , that the stuff we all things we are, we doing is for, to deliver something that's useful. And to be useful, I need to naturally, we need you to confront these values. We need to check how good it is, how bad it is, how, how it's performed along, uh, along the time.

Interviewee

Yes. Yeah, for sure.

Interviewer 1

And what is the feedback you guys usually?

Interviewee

Well, this, it's the, usually, it depends, uh, if you have a static system. Static system, like I say, well, the features inside data does not change too much, uh, along the time, uh, it's easier. The problem is when this, this, this, Changes, uh, changes periodically or randomly or anything.

Interviewee

And naturally what happens? You need to retrain your mother again. Uh uh, no. Uh, Naturally. It depends. The, the kind of problem. Yes. Uh, sometimes you, the most you can get for a, the most prestig you can get from a problem is 60%. And that's good for financial market, for example. It's common. Okay. It's common.

Interviewee

60% of precision. They, they think it's okay. It's good. It's good for, because of the nature, the nature, the nature of financial market, they. Stochastic, um, social all the time. It's change. Uh, the, well, sometimes people are more optimistic than yes, they are more to today. And because of that reflects the, their action inside the financial mark because of that, uh, the data you, you.

Interviewee

Yesterday are slightly different from the data you get yesterday. Well, is the, the, the nature of the problem? For sure. For sure. Yes. Uh, uh, uh, this nature of the problem you are dealing with, uh, this, this, this, this responsible to, to. This is the main responsible to, to the, the quality thing. This is the main responsible.

Interviewer 1

I see. I see. Thank you. And have you encountered any other quality issue during the evaluation of your model?

Interviewee

Uh,

Interviewee

say, if I remember. Well, usually the problems I'm dealing with, Usually No, not, no, not because we have already this second learn stuff. They are really robust. So you can compare to, to sensor floor, to car or to . You see, they, they are robust, so usually no. This kind of problem No. Regarding the model or the, the model or the, the algorithm in, in itself, no.

Interviewee

I.

Interviewer 1

Okay. Uh, perfect. And, um, so my next question is, uh, what are the main quality issues you have encountered during the deployment and the maintenance of a machine learning support system? .

Interviewee

Yeah, yeah, yeah. Yes. Well, mainly, yes. I, I'm dealing with Phil right now. I, I just wrote a, a algorithm to, to, to. Uh, using, yeah, this, this is linear, a linear model, this a linear model, and I, I, they, the goal is this precision of this model is good.

Interviewee

The problem right now is how I provide this model to, to be used to for, for the public in general. Well, the main problem I see, uh, I was trying to use, uh, today, today the, the most popular stuff to put it in the internet. Naturally. The problem is, the problem is what the technology you use to put this at the internet with security, scalability, and, and.

Interviewee

To in cheapy way, in a cheapy way also. Um, you can, you can, for example, in the cloud am uh, Amazon, AWS or Google recipe or wherever the probably fit is the cost. Uh, you, you don't know how many people will, you will be using the system and they don't have any kind of, Any kind of, of, uh, stuff to prevent so many people, so many people to use this system at the same time.

MLP-costly-platforms

Interviewee

And every time these people use, they discuss you money. I. The best they, the best they provide is something that show you how many, how many money are you, right? Real time, money spending. That's the best, the best, the best tool. They, they are, they are. They're providing right now, and this is dangerous.

MLP-costly-platforms

Interviewee

Imagine you have $1,000, for example, like chupa into your project, and this, this value can be depleted in seconds. Yes. And uh, and you tackle this. I was trying to use JavaScript with Python, but I'm not a IT guy and because of that I'm struggling to, to construct the system and what the environment, the Python environment are.

MLP-costly-platforms

Interviewee

Our, our pro our they are offering right now is to use pipeline, uh, Plotly, extremely. This kind of. By now, it's okay, but I, I have doubts in the future how it's it, how it would work, uh, Chabe, it's, this is free, but I think won't be in the, in the future, and I need to tackle this problem right now.

Interviewer 1

Yes. Yes. I see.

Interviewer 1

And I, I think you also mentioned, maybe I'm wrong, um, privacy or security earlier on. Yes. Uh, does it ring any bell? Yeah. Okay. Um, yeah. Security. Yeah, yeah, yeah,

Interviewee

yeah. Go for it. Yeah. The, the privacy security is, are people or robot using your system? The main, the, the main problem. The main problem. I think it's not easy to tackle this problem.

Interviewee

It's not, And the second one is, how do you prevent this robot of the people or wherever, how they inject problems into your system? Because you are providing information for internet. If you are using cloud for this, no problem, the cloud is responsible for that. And you don't, you, you have almost no worries.

Interviewee

But for, uh, ipro, uh, for I promises, uh, on promises system, yeah. This this's your problem. You have to deal with it

Interviewer 1

security person. I see. Sure. Yeah. Super. Thank you. Okay. Um. And I'm gonna name a few quality aspects, and basically if you have experience with any of them, uh, feel free to mention them. Okay? Uh, so fairness, robustness, explainability, scalability, privacy.

Interviewer 1

And these two, we already talk about them. And security.

Interviewee

Security we talk robustness. Yes, because the robustness depends on the quality of the data. We are collect. and you, and you have the system to check the quality of this data. It's, it's, it is according to the, the previous, the previous one you had in nature for, uh, at least. And, uh, and the, and this dispr in another problem is this, this thing, like I said, this, I betray waiting and.

Interviewee

This is a, I don't know how to categorize this into the, into the, this category you say, but today for me, it's a bigger problem. And, and you, you see this problem spread around the na all the industries, even financial markets, even in engineering and brokers and, and pharmaceutical. Pharmaceuticals less because they, they, they are, they are robust.

Interviewee

They, they, they need to, they need to shack a lot of things. But, uh, well, uh, ato, my god, this is most common. I had some, I had some, some, some friends here they are dealing with this, this kind of thing. Collecting, collecting. Information from the court here in Brazil and they, they, they have some models.

Interviewee

Mainly with natural language process, uh, processing this kind of things. And they turn this, this speech into text and they need to check this text to see this with this phrase is good or not for the business. And. You've , imagine how, how complex is, is it, and the model to, to do this is a high complex as well.

Interviewee

And my God, this, this data set is huge, huge. It's horrible to, hard to, to hard to hard to filter what's, what is good, what's not. Hard to filter. What is, what, what phrases you can take as, as, um, uh, what's this phrase is good or not can be, it's a joke or not this kind of thing. , you can imagine how hard it is, but yes, yes, it's a big problem.

Interviewer 1

I. Thank you. And my final question is, um, in your opinion, what is the most pressing quality issue researchers should try to solve?

Interviewee

For sure, by far the quality of didn't say the, the dataset. This quality, the, the data. The way you collect data, the way you engineer this later. And, uh, the way this you transform categorical data into, uh, into numbers to, for your, your model, for your algorithm to prevent this, this waiting, arbitrary waiting problem, like I said to you.

last-question-answers

Interviewee

And, uh, by far, by far this, this dataset is our main problem for, by far, like you said in English, trashing, trash out, no , nothing's different of it, nothing all the time. This is the main problem for engineers. The same problem. You, you can construct the, the best model, the world, but the quality of data is bad.

last-question-answers

Interviewee

The qual, the, the quality of what the, the output was would, would be bad with us. So yeah, there's no, there's no way to, to not tackle this, to not handle this problem. That, that's a very

Interviewer 1

good point. Alright, so I mean, for us, uh, it's finished. We asked all our, all of our question and, uh, I'd like to thank you for attending this interview.

Interviewer 1

It was really interest. And I think what you mentioned will be really useful for this study. Thank you.

Interviewee

Thank you very much. For sure. Thanks again. And you can keep in touch. No problem. You can, you can reach me all the time. You'll have my, you, you guys have my, my mail. If you want to ask anything else, I'm free to.

Interviewee

I'm free to help you or any issues you. Thank you, you very much and if I can help

Interviewer 1

Yes, it, it, it's kind of you. Thank you.

Interviewee

Okay. Thank you. Thank you very much. All right. So, uh, have a good day. Thank you guys. Thank you. You bye. You too. I hope I have hope you and your, your, your jobs.

Interviewer 1

Thank you. Thank you.

Interviewee

Wish you all the best.

Interviewee

Okay. Thank you very much.

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