0:42

Alexey

**This week, we'll talk about making machine learning interpretable. We have a very special guest today, Christoph. Christoph has a PhD in interpretable machine learning from the Ludwig Maximilian University of Munich. He worked for a while in both academia and industry and now, he is a full-time tech writer. He has written many technical books. If you're watching this video, there is a bookshelf behind Christoph. Yeah, with all these four books at the top – we'll probably talk about some of them today. It's a great pleasure to have you, Christoph. Welcome to the show.**

1:23

Christoph

Yeah, thank you for the invitation.

1:25

Alexey

**The questions for today, as always, were by Johannah Bayer. Thanks, Johannah, for your help.**

# Christoph’s background

1:32

Alexey

**Before we start with our main topic, let's talk about your background. Can you tell us about your career journey so far?**

1:40

Christoph

Yeah. My background is in different phases. Phase one was the classic statistician education, I would say – a Bachelor’s in statistics, Master’s and statistics – very classic, hypothesis testing and stuff like this. In the end, I was already a bit drawn to machine learning – I did Kaggle competitions and stuff. Right after I finished my Master’s, I went to industry.

I did one year at a fintech startup, where I worked as a data scientist, then two years as a more classic statistician, where I worked in medical research. During that time, I already started writing the book Interpretable Machine Learning, because I had a part-time position. I just had different side projects and one of them was the book. That's a bit ironic, because after my Master's thesis, (I really hated my master's thesis – the writing process. So it's quite ironic that I picked up writing again. And as I was

2:44

Alexey

**It’s a different kind of writing, right?**

2:45

Christoph

It's very different, yeah. I think that was the reason why it worked for me. When I was writing the book, it kind of took the fear out of writing again for me, because the main reason I also didn't do a PhD was, “Oh, I have to write papers. I hate writing. I hated my Master thesis, so I should not do a PhD.” But when I started the book, I realized I could… [cross-talk]

3:09

Alexey

**So you had a fear of writing, yet you thought, “Okay… Let's write a book. Let's see how it goes.”**

3:15

Christoph

Yeah, because… I don't know. For the Master's thesis, I was very stressed out. It was not only about writing… Maybe it wasn't fear of writing, but I hated the process. All

3:28

Alexey

**Experiments – all that, right? And then getting feedback from your… At least this was how it was for me. Doing all these experiments, showing it to my supervisor, and then he would say, “Hmm… Something doesn't seem right. Maybe you need another half a year.” Right?**

3:45

Christoph

[chuckles] Yeah. And I also didn't feel so free when writing the Master's thesis. But when I started the book, I wrote it in the open, got feedback, and could just write how I felt. I [could] put in some jokes and not hide behind math or anything, just say, “Okay, this is like this.” I use analogies and stuff like that. So I enjoyed this part very much, which then led me to do a PhD. So [I went] back to academia. Then I, again, learned to hate writing. [chuckles] So I really hated writing papers. But still, I finished my PhD. This was one year ago. Then I actually tried a postdoc job for a few months and another job in industry for a few months, and it didn't feel right, kind of. I quit both very quickly.

Then I thought back, “Okay, what's left? What can I do?” Because I now have quit all these things. So I looked back and the only constant thing was writing for me, which I always enjoyed. Then I decided, “Yeah, why not do this full time?” I already had my first book, so I had a little bit of income from that book and already a somewhat established reader base and some experience. So I jumped into this, and I've been doing this now for a year and a half and yeah – it has worked out so far. [chuckles]

5:16

Alexey

**Well, I know that Munich is not the cheapest city. So apparently, it's working out pretty well for you, right?**

5:23

Christoph

Well, I'm earning less than I did in my PhD. Usually people tend to climb up in their career in terms of how much they earn. For me, each career step was a little less money, [chuckles] which is a bit funny, but I have so much freedom that I don't regret it. But when I started, my first job was in Switzerland. My second job was a part-time job so it was a little bit less. Then back to Munich, a little bit less. [chuckles] And now, with my writing, again, a little bit less than a PhD. But I think a full-time PhD is not… I mean, it could earn more, of course, now if I would go for a regular data scientist role. But I enjoy the freedom a lot that I have now as a writer.

6:09

Alexey

**So You can just go to one of the lakes in the Alps. [Christoph chuckles and agrees]. Well, maybe it's a bit cold for that now. But still, you could do this if you wanted. Right?**

6:22

Christoph

Yeah. And I can do sports in the middle of the day or start cooking earlier – I love cooking, for example. But still, I think sometimes about writing and so make progress there at the same time.

# Kaggle and other competitions

6:37

Alexey

**Do you remember what your first Kaggle competition was?**

6:40

Christoph

The first one? This was during my Master’s, and I didn't yet know much machine learning. My first models were a linear regression model and a generalized additive model. I think it was something with diabetes, but I’m not sure entirely what the topic was. But I remembered that I was really, really bad at the competition. [laughs] And this was also like this for me. Because in theory, I mean, I knew how to model data, how to build a prediction model – but just from a statistician point of view, not from a machine learning point of view. My first thoughts were something like, “Okay, this should be a generalized additive model with these features.” I didn't even have a validation set up on my side and haven't fully understood…

7:30

Alexey

**This is not what you do in statistics, right? [Christoph agrees] I remember my first Kaggle competition – I did super bad. Because I thought that with all this knowledge I had from my courses, I would just ace it. The reality was the complete opposite. Very interesting.**

# How Christoph became interested in interpretable machine learning

7:50

Alexey

**How did you become interested in interpretable machine learning? Is it something you were doing during your Master’s? Or is it something you needed to do at your work (in that tech startup)? What led you to it?**

8:03

Christoph

For me, it was quite natural, I think. Because as a statistician, you always think about the data generating process. And when you build modelы, they're usually interpretable. Еhis is the mindset I came from, “Okay, let's understand the data. Let's build an understandable model.” And when I transitioned more into machine learning, this was kind of lost – not fully, because there were things like random forests, which would have this built-in feature importance and stuff like this. But it felt like, “Okay, there are parts missing.” Then, during my second job in Switzerland, I had these side projects and during one day off, I read a line paper, which is about an interpretability method that fits linear models to explain individual predictions of more complex models.

This was quite fascinating to me, like, “Oh, well, we could have a really complex model and then I'll still have a chance to understand what's going on inside the model.” I wanted to learn more, but I couldn't learn more because there was not that much material online. That was my motivation to write a book about this topic. I was quite lucky with the timing as well, because there was nothing on the market yet [about that].

# Interpretability vs Accuracy

9:27

Alexey

**Speaking of Kaggle – in Kaggle, they don't really care about the explainability/interpretability. All they care about is predictions. Right? You care about all these predictions and it doesn't matter how exactly the model is making these predictions. Sometimes it works. At work, you also have similar situations – like churn/not churn. Maybe if it's a lot of users, you don't really go and dig deeper into each individual user when you want to understand why a particular decision happened.**

**As long as the business metrics are satisfied, everyone is happy, then you just stack more layers – or, I don't know, do something. You don't necessarily go and try to figure out why the model was wrong for this particular customer. I imagine there are many, many cases, especially in medical research, where you do have to care about this.**

10:28

Christoph

I think I don't fully agree with the Kaggle case because in Kaggle, of course, what gets you the money is to be right with your predictions. I think interpretability can help you even with Kaggle (go to the back a little). A really stupid example, but I'm doing a competition at the moment. When I created the futures, I kind of had a target leakage because a version of my target, I put into my features. But I also had interpretability at the end of the model, to look at the feature importance using SHAP and immediately… So first of all, I saw the prediction error go down by a lot.

I was super happy. But then a little voice in my head said, “This is not true. Don't be happy. There's something wrong with your model.” And it turned out to be wrong. I super quickly found the bug. For the feature importance ranking, this one feature bubbled to the top. Now, you won’t need feature importance for this, of course, because you could check your column names, for example. But if you have a large feature set, then it's quite tedious to look through all of them. For these interpretation methods, if you have a similar mistake, like target leakage it just bubbles the most important feature to the top. So I would say interpretability is useful for debugging your model.

# Christoph’s current competition engagement

11:59

Alexey

**So you still take part in competitions?**

12:02

Christoph

It's been some time, yeah. It's my first competition in many years. It's not on Kaggle. It's on… What is it called? Data-driven, I think. It's for forecasting water volume through rivers.

12:22

Alexey

**Yes. Speaking of flexibility. We have a… I don’t know if I can say “job,” but with this type of activity, you have the freedom of doing other things, like taking part in competitions. Okay. So you realized that this is an important topic, there wasn’t much literature or information available about that, and you thought, “Hmm… I don't like writing, but maybe I should try doing it one more time.” Right?**

12:57

Christoph

Yeah. I just started to… I don't know my thought process. I was just like, “Wow, this would be a cool opportunity.” At this point, my Master thesis was… I think two years behind me – maybe I just forgot how it felt. [chuckles] But also, it was kind of stress-free, because I could just write – it was a side project for me, so no stakes involved at this point. I had no pressure of having to do the writing so well that I get a Master's degree or get a paper published or anything. I could just write how I felt it to be the best way. So this kind of brought the fun back to me for writing stuff, because I really enjoy understanding difficult things and writing them down in a simpler way. Writing is one way you can express this.

# How Christoph chooses topics for books

13:57

Alexey

**So the best way to learn something is to teach something, or at least write about this. Right?**

14:01

Christoph

Yeah, it's a bit like this. It's also a problem because sometimes, when I write my newsletter, or when I pick a new book – when I already know the topic, I feel like there's not that much value in it, which is obviously wrong because, at some point, I didn't know these things. At this point in the past, I would have loved to read about them. But for me, it feels like if it's too easy to write about, it's not worth writing about it. This is obviously wrong because if enough other people see value from your writing, then your writing is valuable. So I kind of mix these up…

14:39

Alexey

**[cross-talk] … if it’s an easy text. Because machine learning is a complex thing and if there is some resource that explains complex things with simple language, then it's good. Right?**

14:53

Christoph

Yeah, it's good. But sometimes I feel like “I already understand this. I shouldn’t write about it.” But sometimes that's maybe the most valuable thing to write about. Because I already have understood it, and maybe it was hard for me to understand it, but if I want to write about something new, I always feel like I also have to learn it right now. But also, one of the reasons why I became a writer was so that I could always learn new things. It's kind of like a dance between these two things – it should be simple, so I can spend not too much time on one thing, but also, I want to write about something that's exciting for me as well.

# Why Christoph started the writing journey with a book

15:33

Alexey

**Why a book? Usually the story I hear from authors (and I also published) is… For me, I first wrote an article, then a bunch of articles, and then I thought, “Okay, maybe I should do something bigger.” But you just jumped into writing a book without really liking to write.**

15:55

Christoph

Yeah, good question. For me, I already did some blogging before, but I always quit after a few months. But I already had a little bit of experience with writing more freely. With the book, it wasn't like, “Hide in my room for two years, and then finish the book and publish it.” I published it as a website – Interpretable Machine Learning is still available for free. I published it chapter by chapter. So it kind of felt like blogging but, of course, it was one unit. However, [this approach] allowed me to get feedback very early on, and to write chapter by chapter. It didn't feel like one huge chunk – like a huge mountain to climb – but little steps. I think if I hadn't gotten any feedback in the beginning, (I shared it on Twitter, for example) I’m not sure if I would have finished the book, or it would have just been like half a book or a much shorter book.

# Self-publishing vs via a publisher

17:07

Alexey

**But also, as we spoke briefly at the beginning, you don't have a publisher – you published on your own, right? [Christoph agrees] This requires a lot of self-discipline. If you don't publish a chapter, in the case with a publisher, the publisher will chase you saying, “Hey, where's another chapter? The due date was half a year ago. So something.” But here, if you're late, you have to chase yourself. So you have to have discipline for that. [Christoph agrees] How do you do that? Is there a secret, or do you just like the topic and writing so much?**

17:48

Christoph

I would say that I'm inherently motivated to do the writing. But sometimes, to force myself to bring everything together, I sometimes use deadlines for myself, as well. Sometimes you just feel like, “Okay, you could write more, or maybe add another chapter.” But sometimes you just have to finish, and a deadline is then important. I actually was approached by publishers – typical technical publishers – and they asked me, “Hey, do you want to publish Interpretable Machine Learning with us or maybe also write a new book on topic XYZ?” And I always research them, how they work, and so on. Then I read about it, and I was like, “No, that's not my style. I don't want to have their strict schedule.” I also heard that sometimes the editors they have are not so good. Then, also, I get a much smaller share of the royalties. And I was like, “No, I'll do this myself.”

# Christoph’s other books

18:58

Alexey

**We have already spoken about the books you have on your bookshelf. Can you tell us what these books are?**

19:04

Christoph

So, I've written four books. One of them is Interpretable Machine Learning, which is my first book. There's also a second edition already. This was already finished during my PhD, and it's about different… Basically, each chapter has one method that allows you to interpret your machine learning model. My second book, which was when I already started doing this full time, is Modeling Mindsets. It's a very small book. It's also quite different from the first one. It's about the different ways to think about modeling your data.

I already mentioned how I, as a statistician, participated in the Kaggle challenge and quickly realized, “Okay, I won't get far with the way I approach modeling.” So things like these inspired me to write this book. My third book is quite different – it's about conformal prediction, which is a method to quantify uncertainty of your models. It's an introduction to this topic. This was purely because I wanted to learn about the topic and so I wrote the book – it’s like an introduction book. This was a very classic case of learning by teaching.

20:17

Alexey

**What’s the name of the book again?**

20:19

Christoph

The name of the book is Introduction to Conformal Prediction with Python.

20:23

Alexey

**Confounded…?**

20:25

Christoph

Conformal. Like “formal”.

# What is conformal prediction?

20:27

Alexey

**Conformal. [Christoph confirms] I don't think I ever heard this term. What does it actually mean?**

20:35

Christoph

Conformal prediction is about… The basic idea is that you have some kind of uncertainty, and your model and conformal prediction is a kind of process – there are many different algorithms – and they conformalize this uncertainty measure so that you can trust your output. For example, you could have a classifier that outputs class probabilities and you could use conformal prediction to say that, instead of outputting one class or just the probabilities, you get a set with some guarantees (that's important part, these guarantees) of probability that has like a 95% chance that there are correct classes within these sets. You can also use conformal prediction for quantile regression, for example, so that the quanta really contain 80% of your data, of your predictions, of the actual outcome.

21:38

Alexey

**We spoke about this example of churn prediction. So here, instead of just saying, “This John Doe will churn with a probability of 80%,” it would have a range (an interval) of how certain we are that this prediction is correct.**

21:55

Christoph

Yeah, that's also one method. Conformal prediction is kind of… For classification, we can kind of put a balance on the probabilities. You don't just give one probability but a range between what you believe… [cross-talk]

22:18

Alexey

**Is it related to… I guess there is some relationship between interpretable AI and this topic. Because to come up with this balance, you kind of need to do similar things, right? Or am I completely wrong?**

22:33

Christoph

So, technically, I will say it's different. Because for conformal prediction, you separate your model… So, let's begin with the similarities. It's a bit similar to these model-agnostic interpretability techniques, in so far that you can first train your model (any model, really – it could be random forest, it could be neural network) and then apply the method afterwards. Because, at least for some classes of conformal prediction, which are called inductive conformal prediction, these can apply to your model after you've trained it. The idea is that you have a calibration step so you can use some output from a model which quantifies the uncertainty and kind of calibrates it the right way. But otherwise, yeah, that's the idea… [cross-talk]

23:26

Alexey

**The methods are completely different, right? Only the way you can apply it. There is a model running in production, so you don't need to actually do anything with that model – you just do something on top.**

23:37

Christoph

Yeah, that’s at least one possibility with inductive confirmed prediction. Yeah.

# Christoph’s book on SHAP

23:44

Alexey

**Interesting. And the fourth book?**

23:49

Christoph

Yeah, that's Interpreting Machine Learning Models with SHAP . It also has Python examples, which my first book doesn't have, about interpretable machine learning. But otherwise, it's kind of taking one of the chapters from Interpret Machine Learning and doing a deep dive into one of the methods, which I think is also the most popular method (SHAP) and also one of the most versatile methods. So I decided it would be cool to have a book on it. This was one of the books where I was like, “Okay, I really have a lot of material already – and I still learned a ton while writing the book, because you realize that you still have many gaps – but still, this was where I kind of took the stuff that I already had and not only learned new things. [chuckles]

24:41

Alexey

**How did you decide that you want to write about this particular method? Did you do some research that showed that this is the most popular one? Did your readers maybe ask you for more examples for more hands-on examples with code? What was your process? How did you arrive at the idea that you want to write this particular book?**

25:03

Christoph

Yeah. For SHAP, I already had some insights as it is hugely popular. So it was kind of market research because I already know from my book, Interpretable Machine Learning, from Google Analytics, I know which chapters are popular. SHAP… I also have a separate SHapley values chapter and they kind of belong together. They are by far the most popular ones. So this was…

25:30

Alexey

**Thay had the most visits, right? Because you can see how many times each chapter was viewed, right?**

25:37

Christoph

Yeah. Also, I think SHAP is not super intuitive. I mean, it has lots of material already, like blog posts, and so on. But I thought that putting this all together in a book would be valuable to many readers. I also find SHAP interesting because there's always this danger – you shouldn't always only rely on one interpretation method. There was already a blog post about it – SHAP is *not* all you need. But if someone would force you to use only one, I would use SHAP, because it's very versatile. I think there's a lot of value if you learn it.

# Explainable AI vs Interpretable AI

26:17

Alexey

**Well, maybe we need to speak about this particular method. But I see that there are quite interesting questions. The first question is from Rosana. The question is, “Is there any difference between Explainable AI and Interpretable AI?”**

26:31

Christoph

Many, many people ask this. I don't think I have a satisfying answer. Some days ago, I posted about an overview of how researchers use this differently. How I use these terms is – I use both interpretable machine learning and explainable AI as keywords. I think that's most useful because there's so much overlap. But people just use both quite interchangeably. So if you want to do a Google search on some topic, I would always use both. I think I make a distinction when I use words – when I say, “I want to *explain* a prediction.” Some people would criticize me now, “You don't explain a prediction,” because it's a very strong word, obviously. But I think it's still fair to use it.

So this would be more for a local understanding of the model. And for global understanding, I use the word “interpreting” more. I know that others like to distinguish between interpretable models – with this, I mean that they are kind of inherently interpretable, like a linear regression model – and explainable AI like for these postdoc things. But I don’t think that's a good distinction, because the boundaries between the two are not super clear. So I'm sorry, but I don't think I have a very satisfying clean answer.

28:08

Alexey

**So not really much difference. The question was about explainable AI and interpretable AI – but you corrected it to “interpretable machine learning,” right?**

28:19

Christoph

To cover both, yeah. Interpretable AI, explainable…

28:19

Alexey

**Is that how they are usually used? Is this how the term is? It’s always “interpretable ML”?**

28:29

Christoph

I'm biased because I named my book this way. So, in my head, it’s always “interpretable ML”, and not “AI”. But you can also use this. I guess the question boils down to the definition between interpretable and explainable – I don't make a big distinction between the two.

28:50

Alexey

**Well, if you think about the words themselves – what's the difference between “explain” and “interpret”? For “interpret,” you interpret it, and you don't necessarily explain it to others. [chuckles]**

29:00

Christoph

Yeah. for me…

29:02

Alexey

**Like “I understand how it works.”**

29:03

Christoph

…”explainable,” for me, is a very strong word. I think that was also why I said interpretable machine learning when I named my book. But I named it quite early on, so a lot of things happened afterwards and the term “Explainable AI” got quite famous as well. At some point, it actually was named Explainable AI (my book) but then I decided against it, and now it's Interpretable Machine Learning”. I think I like “interpretability” a little bit more, because “explainability” is very strong. To explain something, you have deeply understood it. But interpretation is like, “Okay, you see something and you have an interpretation, because you don't *fully* understand it, but just parts.” You kind of derive with your insights, which is an interpretation of what you see. So that's why I like the term more

# Working alone vs with other people

30:00

Alexey

**Okay. Now you’ve been working as a technical book writer for a year, right? Do you feel lonely when you just write by yourself? Do you miss colleagues?**

30:19

Christoph

Sometimes, yes. I like working by myself a lot. But sometimes I'm like, “It would be nice to work in a team for a day or so.” The problem is, you can’t really do this on and off. I mean, you can, but then you're tied to a work project and… I just would like to have this some day. I mean, it's not fully lonely. I'm talking to you right now, for example. You can do other things like podcasts and talk to people.

Also, I am writing a book at the moment with Timo Freiesleben, who's a former colleague of mine. We’re writing this book together – it's my first collaborative book. I also have a little team right now, at least for this one project. So it's not only me sitting alone in my room and tinkering and writing. [chuckles]

31:18

Alexey

**I guess writing with somebody else not only makes the process… It kind of forces you to finish earlier, because half of the content is written by you and half of the content is written by your co-author, and then you review each other. So, you kind of deliver the book faster, right?**

31:41

Christoph

I would say it's slower because my co-author has a full-time job. Like to. He’s a postdoc, so he can’t dedicate as much time as I can. So if I would write it alone, I think it would be faster. But it's fine the way it is. The topic is supervised machine learning and science. It's about how supervised machine learning is used, like for doing research and stuff – so, kind of, applied machine learning. The book has two parts. It's about the justification of why it's okay to do it, or whether it's okay to use machine learning because there's some problems, of course, when you don't fully understand the model and so on.

But also there are puzzle pieces that if you use machine learning in science or in research, what do you actually need to make it work? Because if you just use it in something like a Kaggle competition, then there's lots of missing pieces, like causality, interpretability, uncertainty, quantification, etc. So we're kind of putting this together and I think it's good that we’re taking the time to write this. We have a half draft now. But we're taking our time to really think this through, because it's a new topic and there’s also a philosophical component to it. I think it's a fun project.

# Christoph’s other engagements and how to stay hands-on

33:07

Alexey

**So you said you would write it faster because your co-author has other commitments. Then you also don't invest as much time as you could potentially because you probably… What I'm trying to say is that you probably still have time [left over]. So are you working on something else in the meantime apart from the book?**

33:27

Christoph

So right now, I'm participating in a challenge – not on Kaggle, but another platform.

33:34

Alexey

**Driven data, right? The one that you mentioned?**

33:38

Christoph

Yeah. It's almost full-time at the moment, since a week or so. But that's the first time I'm doing this since I've been writing. I felt like I need a hands-on project again because sometimes it can feel like a disconnect, if you always write about using machine learning and so on. I do have lots of experience from earlier on. But the last one and a half years, I've only been writing and doing no data analysis, really. Then it feels like I’m a bit out of touch maybe. So it was really good to do. It still feels really good to do a hands-on project with all the problems with the data and how to model it and so on.

34:20

Alexey

**It's nice that you brought this up, because I also… I used to work as a data scientist full-time and now I don't. What I'm afraid of is that I'll lose touch with reality. Because if somebody asks me about something, then my answer would be based on my experience that I got at work – but I'm no longer working, so this answer might be outdated. The longer I don't work as a data scientist, the more outdated my answers become, right? [Christoph agrees]**

**So how do I stay hands-on? This is what you’re doing, right? You’re taking part in competitions and you make sure that the methods you teach (the things you write about) are still applicable because you can apply them in these real life projects.**

35:13

Christoph

Yeah. But I feel like… I didn't expect this, but the writing helps me to be better in the competition. It's because I'm now much more structured. I actually have a little logbook where I write about when I'm stuck, for example, or when I have a breakthrough. I didn't expect this, but it helps me a lot to keep my thoughts in order because in a machine learning competition, you have so many directions that you could follow. Very often, you try something and it doesn't work.

I still get frustrated when it happens, especially if you invest a whole day. But if you can write it down, you can process it, I think, much faster and be like, “Yeah, this was an experiment. The experiment showed, for example, this weather data didn't help anything with the prediction or improving the performance. So I’ll move on to the next experiment.” To keep myself focused on what I'm doing, it helps me so much that I have this habit of writing. It’s quite unexpected.

# Keeping a logbook

36:21

Alexey

**This is so cool. I want to learn more. How exactly do you structure this? What kind of tools do you use? Is it just a Google Doc or something else?**

36:31

Christoph

You mean, especially for this competition?

36:33

Alexey

**Yeah, for this competition, and in general. Let's say, I’m taking part in a competition and I want to follow your approach. I want to have this logbook. How should I do that?**

36:43

Christoph

It's extremely messy. I'm doing this for the first time. It kind of works, but it's not like I have a fully-fleshed out process on how I do this.

36:52

Alexey

**So it’s not like when you finish the competition, you’ll have a book that you can publish, right?**

36:55

Christoph

Yeah, I actually thought about that. Because there's communication, there's prize money involved in this project (in this competition). Mostly, of course, it’s performance-based, but there's also a competition part or a prize part about explainability and communication.

37:13

Alexey

**Ah, I see why you’re taking part in that competition.**

37:15

Christoph

Yeah, that was a bonus. Also, the topic was interesting to me. It's about predicting the water flow of rivers, and I like all these ecology data tasks. One thought that I have is that I could also use this journey to kind of write a narrative for how I approached the entire challenge. I try to write down every day what I’m working on in the challenge – a few thoughts at least.

37:45

Alexey

**Is it like a diary? Is it like a journal where each time before going to bed you write “Dear Diary, today this is what I tried.”?**

37:55

Christoph

Yeah, it's a bit like this. I use Obsidian.

38:00

Alexey

**It’s electronic, I assume – not paper.**

38:03

Christoph

It's just very messy. The idea is that it shouldn't take a lot of effort to do it, so I write very hastily, very short sentences and just bullet points. As a note-taking tool, I use Obsidian. Also for other things. For this challenge, I have no one file, and I just dump every thought in there. Also, this logbook of what I do. This helps me keep focused on the project.

38:38

Alexey

**I guess you have these bullet points in case you want to go back to them and turn them into a proper text. You have all the information there. Right?**

38:48

Christoph

Yeah. One reason that I have this is because I just wouldn't remember, “On day seven I struggled to make a good submission,” or whatever. The other reason is because, just writing itself –even if I don't ever look at it again, if I don't ever touch it again – it helps me to put into words or to summarize what I did the entire day. It's super frustrating if you just don't think about what you did all day. But if you write it down like two sentences, “Today, I've finished this part and I realized that using weather data doesn't help,” then it already helps me to keep focused on what I’m going to do the next day and to not get frustrated that maybe I didn't make as much progress because often you do more than you think, because you've tried many things. At least that's how I feel.

39:40

Alexey

**I was going to ask, “Why would somebody even do that? Why would somebody even write it ‘On day seven, I struggled to make a submission.’” Because it's not useful. But when you say that it puts things in perspective – it's not like on day seven you slacked and didn't do anything. [Christoph agrees]There were some problems and you needed to work on solving these problems. So now you know that “Okay, I actually worked on something.” It keeps you motivated, right?**

40:10

Christoph

Yeah. It's hard to put into words, but sometimes I struggle with something, but I don't formulate what the problem is in my head. And then sometimes I get distracted and go to Twitter or do something else. But writing down what the essence of my problem is – for example, in this particular challenge, the problem (or the goal) is to predict the volume in a season. That's the total volume over a few months. But you have to make the predictions at different points in time. Some points, or some dates, are actually already within the season – so you have some partial data (some partial waterflow) already.

I really struggled with how to put this into my model – this knowledge – because, obviously, if you already know that some amount has already flown through the river, then you can adapt your prediction. This was kind of my problem – I was thinking about it for a few days, actually. But this helped me, to put this into writing, like, “Hey, this is what I have to achieve. This is the problem I have to solve – the most important one.” It's almost psychological, maybe [chuckles] a mental process to help focus.

41:32

Alexey

**How much time do you spend on this logbook?**

41:35

Christoph

Very little. In the beginning, I was just writing down parts, “Okay, to understand the evaluation metric, I just put all the information in.” And things like, “Okay, these are the problems about the data.” I put that in one section. In the beginning, I spent maybe a little more. But this was while I was reading the description of the challenge – I summarized the most important parts for me. There may be one or two hours there. Now I just spend five or ten minutes a day. It's very short. I just wrote a few lines about what I did today.

42:16

Alexey

**It’s like a diary, except it's for the competition.**

42:18

Christoph

Just for the competition, yeah.

# Does one have to be an expert on the topic to write a book about it?

42:21

Alexey

**There is a question I really like, and I think we talked about that. The question is, “Do you have to be an expert in the field to start writing about it?”**

42:32

Christoph

It helps. [chuckles] Or maybe it doesn't. Maybe it doesn't.

42:38

Alexey

**I can see why it’s not always…**

42:41

Christoph

You can make the argument both ways. If you're not an expert, then you have the same obstacles as your future readers. That means that when you write the book, you can write about these obstacles. These misunderstandings that you had in the beginning about the technique you're writing about or the algorithms you're writing about. But it may also be a very tough process for you because you have to learn and write at the same time.

43:08

Alexey

**And you also need to make sure that what you’re writing about is actually correct. Right?**

43:12

Christoph

Exactly. Yeah.

43:13

Alexey

**It’s like you misunderstood something, but you thought this is the right way and you write about this, and half a year later turns out that it's not and you have to re-do the whole thing.**

43:24

Christoph

It’s also a question if you already know something. Let's say you want to write a book about machine learning, but you don't have the basics. Going into machine learning, for example, was much easier for me because I studied statistics – a lot of the math behind machine learning and working with data was something I already knew. So you should also think about where you should start from. For example, conforming prediction was new for me – the topic itself. But, of course, many of the concepts are new, even if I didn't know conformal prediction before. So I would say, the very best thing would be to be an expert, but still remember all your struggles and misunderstandings.

44:15

Alexey

**So you need to go down to the level of the reader, right? For example, cross validation – for you it’s just waking up in the middle of the night, with your eyes blindfolded, and you can do cross validation, right? For somebody who is new to that, they might not even know why it’s an important thing – how it works, why we need it, and what the gotchas are.**

44:36

Christoph

If you keep a logbook while learning then you could also remember your struggles. [chuckles] So I think this would be the ideal setup – being an expert, but still remembering what it was to be a beginner.

# Writing in the open and other feedback gathering methods

44:51

Alexey

**Or having a beginner who you can experiment with – someone who could say, “Hey, I don't understand this thing.” You can see where they struggle, right? [Christoph agrees] I think this is what you do with your book too, at least we talked about that for your first book. Because you wrote in the open, so everyone could see and comment and say where they don't understand. You could then take this feedback and incorporate it into the book. Do you still write in the open?**

45:22

Christoph

It’s quite different between the books, I would say. The first book was fully open and I wrote it chapter by chapter. The next books were closed – most of the books, I guess. For Modeling Mindsets, I asked friends and colleagues to read individual chapters. And then, later on, for the Conformal Prediction book and the SHAP book, I actually asked readers if they want to be test readers. This worked quite well. Because then I already got people from the audience.

46:00

Alexey

**How did you ask readers? You have a newsletter, right? [Christoph agrees] So in the newsletter you can write, “Hey, I'm writing a book about this. Do you want to be beta users (or test readers)?” Right?**

46:10

Christoph

Exactly. And then people say, “Hey, yes.” And then I share a Google Drive version of the book. So far, this worked really, really well. I got extremely good feedback from this process. I'm very grateful to everyone who reads my books before they are finished. This worked quite well for me.

46:36

Alexey

**Would it be problematic if you have too much feedback? If there are too many people who want to read your books?**

46:41

Christoph

I once had this problem with the SHAP book, where I invited a lot of beta readers. I think one problem is, of course, if you have a very problematic part, then all the readers will comment on this part. It's also not so good that you give this version to a lot of people. One lesson for me was to keep the list of readers small. You shouldn't have too many beta readers, I think. Rather, you should have more phases where you invite a few readers, incorporate the feedback, and have a new version of new readers for this new version. That's what worked best for me so far. But, of course, it prolongs the process of how long it takes until you can publish.

47:33

Alexey

**There's also the part if you don't agree with the feedback – what do you do then?**

47:37

Christoph

If you don't agree, you don't have to use it. I mean, that's the nature of feedback.

47:42

Alexey

**But then the person might feel offended, like, “Okay, he doesn't want to listen to my feedback. What am I doing here?**

47:48

Christoph

I haven’t had such cases so far. I think you should always double-check the feedback, in the sense of, “Maybe I'm wrong.” I always come from this angle, “Probably I'm wrong.” Because the reader is the reader – in the end, I’m writing for the reader. But sometimes, you can also have the case (one that I already had) where one reader says, “Hey, you should have a simple example in the beginning,” and the other reader says “Hey, why is there an example already? I haven't really understood what the method is about, so this should be later in the chapter.” So there can also be different tastes or different opinions and you just have to make the decision on what you think is best.

# Advice for those who want to be technical writers

48:36

Alexey

**And probably on what feels more natural to you personally. [Christoph agrees] I see. Another question I really like is, “Do you have any advice for somebody who is looking to start their journey as a technical author?”**

48:52

Christoph

Yeah. Don't start with a book. [chuckles] No, that's actually what I did. But I would say start with small units. This could be posts (like a blog post) or maybe even smaller – something on social media, LinkedIn, Twitter, Mastodon, whatever – and just start writing. For posts, you can…

49:17

Alexey

**Just start writing, right? That’s what… [chuckles]**

49:19

Christoph

Yeah, just start writing, kind of. [chuckles]

49:22

Alexey

**Just do it. [chuckles]**

49:24

Christoph

Yeah, because I think the biggest issue is when you write but never publish – this is frustrating. You don't get feedback. I think the worst that can happen is, if you publish but nobody reads it. But that's also the case if you never publish – nobody will read it for sure. I mean you can write for yourself, but then it's more like a diary or something. But I think just getting started is the most important thing. It sounds so like… dumb. But yeah – make it small and publish it.

50:00

Alexey

**Maybe also the question is, “How can I be a full-time technical author? What are my steps?”**

50:08

Christoph

Hmm… Yeah, I don't know. I haven't studied many people who do this full-time. Actually, I think most people are just doing this on the side. Maybe they have a successful career, and they write one book or two on the side. Technical writing is something… I wouldn't recommend it, if you haven't done anything like this before – to just quit your job and start it. I think that's a bad idea. Because this is something you can actually build up to slowly.

50:41

Alexey

**That’s pretty much about anything, right? If you haven't done something like that, don't quit your job and start doing it.**

50:44

Christoph

I mean, if you have enough financial security, go for it. But the thing is if you have a job until today, and tomorrow, you say, “Okay, I’ll start technical writing.” Even if you’re good at it and even if you’re fast – your first income will come a few months later. Many months later, actually, because first you have to write the book, then you have to publish it, and even after publishing it, you have to wait like two months before… And even if you make sales on day one, you have to wait two months until you get the money from the sales.

51:13

Alexey

**How much time does it take to write a book?**

51:17

Christoph

That’s really difficult to say. I would say at least a few months. It can take many years, depending on how much time you spend with the book and how many people are involved, and if there's a long editorial process behind it. For me, I would say like, half a year or so.

51:37

Alexey

**Okay. Half a year of working and then after you publish, maybe in three months, you'll start getting paid.**

51:46

Christoph

But also not full-time working, because I think you can’t work full-time on a book. I mean, I can’t write eight hours a day. You go crazy. So I do a lot of other things like, now, the challenge, obviously, but also things like posting on social media, newsletter writing. These are things if you want to be self-published, and you also need some means of distribution and reaching people. Even if you're not a self-publisher, you often have to do your marketing yourself.

52:19

Alexey

**Yeah. I published the book with Manning. It took two years. And then it feels like… I might be wrong, but it feels like they just want me to do all the marketing. [Christoph agrees] It’s like, “Why do I need you guys?” [chuckles] Right?**

52:34

Christoph

Yeah, that's something I didn't expect in the beginning. I researched this a lot. I think it might help if you have a name tag on it from a bigger publisher – then it's in their library (in their shop). Because of the stamp of the publisher, I think it can help. Because, for me, people just have to trust in me, kind of, because there was no decision saying, “Okay…” [cross-talk]

52:59

Alexey

**You do this by already having an audience of followers. From the beginning, you said you were writing in the open, so people could follow you and could see what you’re writing about. So you already got some trust from them, right? Then, by the time you finish the book, they thought, “Okay, this is such a cool book, and we know Christoph, so we want to buy from him.”**

53:22

Christoph

Yeah. I think this supplements [the marketing] because, I mean… I didn't have any credibility in the beginning. But I wrote the book in the open, so people could just see the book, and later on decide whether they want to buy it in paperback form, for example. There wasn't a big trust issue or anything, because they could just read it beforehand.

# Self-publishing tools

53:49

Alexey

**What do you use for publishing? Because you also have physical copies. I know there are websites where you can sell digital products – PDFs, videos, whatever. But you have physical books.**

54:05

Christoph

The first part where I sell is on Leanpub, which is PDF and eBooks. It’s just a digital bookstore. Then there's print-on-demand services, where you basically also upload a PDF file, a cover file, and put on all the details – and then you can self-publish there.

54:26

Alexey

**Do they also handle the logistics?**

54:29

Christoph

Not at all.

54:30

Alexey

**So they send you a truck with books? [Christoph chuckles] They offload it in your yard and then you do whatever you want with this?**

54:37

Christoph

No, that's fortunately not the way. I mean, you could do it this way. It would be much more complicated and you would have to do all the packaging and sending out books. I guess I would earn a bigger percentage per book, but…

54:51

Alexey

**But then you wouldn’t have a life – you would just do this all the time.**

54:56

Christoph

Yeah, I'm using Amazon KDP, but there are others as well. That's a kind of a print-on-demand service. You upload your book there and they sell it for you. If you buy a book of mine, then they will just print it and send it to you. So there's no pre-printing or having my garage full of books involved.

55:21

Alexey

**So when you do it through Amazon, the book appears on all Amazon stores, right? [Christoph agrees] In the US, in Germany, in France… [Christoph agrees] Oh, that’s cool.**

55:30

Christoph

Yeah, you can actually pick the marketplaces where you want it to appear. You can even pick expanded distribution so it gets to the other bookstores as well – the bigger ones at least, like Barnes and Noble and stuff.

55:45

Alexey

**Ah, okay. Even that. People can even see your books in a bookstore.**

55:50

Christoph

Yeah, so it's a good time to be a self-publisher, because there's lots of these platforms and tools available. I actually had some doubts whether it's a smart choice because of ChatGPT and stuff, when it came out. But I’m more relaxed now. I mean, there are lots of books that are auto-generated by large language models, but I still say that it's a good time to be a writer.

# Finding Christoph online

56:16

Alexey

**Interesting. You said that people can find your books on Leanpub and Amazon, but do you have something like a homepage with all the books and links and all that?**

56:25

Christoph

Yeah, christophmolnar.com/books is just a list of my books.

56:34

Alexey

**It's probably a good idea to follow you on LinkedIn and Twitter, right?**

56:37

Christoph

Yeah, that's always a good idea. Of course. [laughs]

56:41

Alexey

**Anything else you wanted to mention that we forgot to talk about?**

56:45

Christoph

Um, no. I just had a lot of fun talking with you. I think it was a very nice conversation.

56:53

Alexey

**Yeah, thanks. Indeed, it was nice. Unfortunately, this is all the time we have for today.**

57:01

Christoph

It was quick. [chuckles]

57:02

Alexey

**That was quick, even though it was actually one hour. It felt pretty short. Yeah. So, thanks for joining us today. And thanks, everyone, for joining us today, for asking questions. Well, have a great week ahead and see you in a week.**