1 Interview

2 **Dan:** Okay. So now I will switch back to the slides and then we'll conclude with with an interview. Okay. So my first question is, what is your current, well, I guess the first question is, do you use PDP and or ICE plots in your work?

3 **Participant:** Oh, yeah, definitely like not only like do we for our own purposes, like we're kind of like even requiring to, it's like part of our kind of like documented model governance and all that so definitely a lot of uses cases.

4 **Dan:** Okay. Do you use both PDP and ICE plots?

5 **Participant:** Yeah, both of those, but mostly honestly, the one-way, I think no real good justification for that other than just like the precedent of what's often been done in our business and quite honestly, I think the two-ways are just um, we often have like a ton of variables in our model and, you know, I mean, I think without the kind of like approaches to sorting like you've developed, they're just like too many two-ways.

6 **Dan:** Okay. And then so for one-way plots, do you use both PDP and ICE plots?

7 **Participant:** Yeah, same sort of use like you have with them, like kind of superimposed. We typically like don't go super deep into the clustering details. So I mean, it's like we look at both, but unless anything's like really wonky. I'd say it's more of the PDP that's like drives the narrative.

8 **Dan:** So I think at one point, you mentioned using ALE plots as well.

9 **Participant:** Yeah, XXX XXXX XXXX XXXXX XX XXXX X XXXXXXX XXX XXXXX, it's something that I like kind of looked around at. But I'd say that's kind of like less common in our general environment.

10 **Dan:** Okay. And now for for the PDP and ICE plots, what tasks do you use them for?

11 **Participant:** Mostly like kind of when you're like building or validating model, just making sure there are no super unexpected or intuitive trends, like to your example that could be causing, you know, could be signs of kind of like overfitting or the model causing spurious patterns or also like learning things that might suggest that it's kind of like also behaving in ways that we wouldn't want it to for like regulatory reasons.

12 **Dan:** And how do you determine which plots to look at?

13 **Participant:** That isn't like totally so much standardized across teams. I think I've seen a lot of kind of like, honestly, like rather naive approaches of just like kind of somebody like literally eyeballing a lot of them or pulling out the like non-monotonic ones. I definitely, I think liked kind of your like, I don't think I don't feel like we have a good solution for that. And a lot of the work I do is more kind of (unintelligible) and as of late has been more in our model governance function kind of like receiving these from like kind of other teams. And like honestly, like it's a little bit of a kitchen sink approach. And you just like kind of take it on faith that they looked at them and pulled out the important ones, which is like honestly a little bit frightening.

14 **Dan:** So can you explain what you mean by kitchen sink approach?

15 **Participant:** Just kind of like making a PDF with like all of them. And then someone saying like, we thought these were important.

16 **Dan:** Okay. And then before you did like model governance work, when you were like the ones like generating the PDP and ICE plots, did you have an approach for determining which ones look at?

17 **Participant:** I mean, I'm not saying this is a good approach, but I think you know often it was in tandem with other like variable importance type measures of like kind of, you just get kind of like printout of what like variables that are having the biggest influence on your model. And then really do a deep dive into like what specifically that influences.

18 **Dan:** Okay. So I guess.

19 **Participant:** Go ahead.

20 **Dan:** So like you would use like other like feature importance measures and then maybe like focus looking...

21 **Participant:** Yeah, but of course like going on to pain points. Like I think the thing that gives me the most heartburn with that is all of that of course is a very, very top down global approach to it. So you're missing a lot of like nuance from the sub pockets. And I think this tool does a lot better job kind of finding the like heterogeneous trends.

22 **Dan:** Great. And so for in your work with ICE plots. Did. Did you do anything in terms like trying to like identify clusters of

ICE of ICE lines or anything in that vein?

23 **Participant:** Honestly, not. Not really.

24 **Dan:** Okay. And I guess apart from. So you mentioned you don't really use two-way PDPs primarily due to like the number of features and, you know, current precedent. Have you seen or have you done other approaches for analyzing interactions between features?

25 **Participant:** Yeah, honestly, like I think not. I think to some extent it's something we have like generally under invested in a little bit, beyond kind of what's needed or like beyond our more specific concerns about kind of like, well, let me say it a different way. I think we often have a slightly more like, more than we should, hypothesis driven approach to things of like we're worried about things that could cause disparate impacts, we're worried about variables like from XXX XXXXXX XXXXXXX that we have, we know maybe like their encodings changed over time or it would have clear like XXXXX impacts, like (unintelligible) XXXXXXXXX XX XXXXXXXXXXXXX XXXXX XXXXXXXXXX XXX XXX XXXXX. So I mean, I think. More than we should, I think we've tended to take more hypothesis driven approach to things where, some features like seem like they could pose risks and just ensuring that those are doing things that seem like they have fair and reasonable intuition behind them.

26 **Dan:** Okay. But I guess apart from like the hypothesis driven like things you look at there isn't like a general approach for like analyzing what are like the big interactions in the model?

27 **Participant:** Not. I mean, the the other tool I guess we tend to use a good amount is almost more like residual analysis. So, you know, kind of like. Yeah. Well, that that's kind of solving a sep- Like, yeah, I think I guess we tend to focus a lot more on identifying model weak spots, which is almost kind of like the reverse case of feature important. You know, I mean, then I think you sometimes kind of back out like. Oh, and that's because like. These features are doing the weird thing in this space, but I think we almost take a more. Starting with your residuals and backing stuff at like residual analysis and backing stuff out as opposed to just like truly like focusing in on just like model mechanics in and of themselves, if that makes sense.

28 **Dan:** And what what would model mechanic in that case mean?

29 **Participant:** Well, I just mean kind of like this where you're really just thinking about like. How do my features affect my outcomes independent of whether or not that's making the right or the wrong predictions sort of thing.

30 **Dan:** Okay. So next we'll move on to asking more about PDPilot in particular. So how would you it supported or did not support your model analysis? Were there any questions you weren't able to answer or any tasks that you weren't able to perform?

31 **Participant:** I think I mean, I thought it was like really easy to learn really is to use and understand and good for it's like primary purpose. I think like kind of an amazing like dream tool for doing this sort of model analysis would somehow fold in and like I say this realizing it substantially out of scope, but, I think somehow like integrating in more of like the aspect maybe in some of the filtering or sorting of like... kind of good versus bad... model performance or you know, I mean, like. I I think if there was a way to kind of like tie kind of the model performance and the model interpretability pieces closer together. That would be really powerful. But that's super handy wavy right now. But I think in terms of kind of like. Definitely when I had a clear question, I was trying to answer. It was really easy to use it to answer that. I think the only other like tiny things I noticed. Like I think I said it in the time was like. With the clusters, I'd love love for those to like be the axes and some of the like distributional plots so I could kind of see like. Within my cluster, how does this feature breakdown instead of like. How are my clusters distributed across the features. (unintelligible) And yeah, I think the. With the housing example, I think I did kind of struggle, but I did keep kind of running into the issue of like. Oh wait, that's really low sample. Oh wait. This is really highlight and this is highly ranked, but that's really low sample. And I think if there is anyway, and again, like I don't know how you'd do it. And I mean, I think there's like already kind of like the right amount of content per page. Or per view. But I think somehow baking in kind of like. Kind of the frequency of some of these things and they I know the histograms do get to that. But I like I think. At least that's something I'd want to like key into a lot more than I was using it. And maybe if that could somehow, if like. I don't know if there'd be an option to like. Let some of the sorting criteria be like partially like weighted by that or something.

32 **Dan:** Okay, that all makes sense. Okay, so let's. So just step through a few of those points. Okay, so you mentioned the thing with the clusters. So I don't remember what exact feature you were looking at. It might have been.

33 **Participant:** I think, yeah, but basically I guess it's sort of just true of any like categorical feature.

34 **Dan:** I think it might have been foundations that you're looking at maybe.

35 **Participant:** Could be, that rings a bell.

36 **Dan:** Okay, but I guess it doesn't matter which particular feature, but I guess your point was that. So like in this plot. We're showing for each feature value, the distribution of clusters, but you would be more interested in seeing for my blue cluster.

How does it distribute by. Across the values. Is that right?

37 **Participant:** Yeah, I think I'd find that easier to just like. Read the profile of the clusters that way. Because it's kind of like conditional on this. The distribution looks like this. And I think the thing I'd want to condition on wouldn't be the clusters.

38 **Dan:** Okay, yeah, that's interesting. You're not you're not the first person to mention that. And even like some past participants when they like they read this plot, they interpret it as like. 100% of the red cluster has major deductions when that's not the right interpretation.

39 **Participant:** Yeah, (unintelligible) I kept almost doing that, yeah, and then I kept being like, no, that there's more red that isn't 100%. But yeah, I kept almost falling into that same trap.

40 **Dan:** Okay. Yeah, that's interesting. That's something that I hadn't considered when when making this. I guess from a visualization perspective, how would you want that to be visualized? Because I guess it seems like it'd be. Like using color twice in like two different ways seems like it might be confusing. So would you want like. Like a separate like bar chart for each cluster?

41 **Participant:** I see your point about not using color in two ways not being ideal, but I don't like. I almost I guess that's almost how you'd have to do it and then just. I don't know if it could just be a different palette. So people didn't get confused. I mean, again, like I know that like. Um. I know optically and it like might look really ugly. But yeah, I guess color is really the easiest thing to like eyeball easily. And I like maybe could like, well, no, that doesn't make any sense. Never mind.

42 **Dan:** Okay and then

43 **Participant:** Because I like the stacked bar idea like. I think it's like the right sort of plot. So I almost don't think there's any other way to do it, but use color.

44 **Dan:** Okay. And then so you mentioned wanting to like better tie in model performance. Do you have any idea of like what that would what that would look like? Because I agree that like with this tool, there's no information really about how the model's performing. So what would you see as like useful ways to integrate that?

45 **Participant:** Maybe it's even almost just in a validation capacity, like for example, um. I don't know if this is using in this plots with the scatter plots. I don't know if you're using test data or training data.

46 **Dan:** So everything's training data here.

47 **Participant:** Okay. But like. Like this is a really good example where you were like, yeah, it looks like the models like super overfitting in this high space and like. Then that that does like raise my mind question in my mind, like. Does that also mean my like. Average mean squared error residuals are also a lot like higher. Or is like, do my mean squared error residuals also trend by this feature? And I could see like, I don't know if like if this were like test data. I don't know if we could if there'd be a sense of either kind of like. Coloring the points scatter plot by the MSE. Or if there'd be a way to kind of like. Plot another line, sort of like the PDP line, but it's like kind of. Average. MSE by some sort of like bins over the skew or. Just kind of some sense of kind of like. Simultaneously like. Is my model finding a major trend. In this, and does that make me happy or does that make me concerned. Because I think I always defaulted almost more to the story telling in my head of like. Oh yeah, this is doing this because of this real world thing, but then I kept having to remind myself like this isn't actual data. All I know is the model's doing this. I don't actually know if that makes me happy or not. And again, like that like I know you can only put so much on the plot. So I'm not saying any of that is reasonable and obviously I've spent like three minutes thinking about it, but just like. Spitballing here.

48 **Dan:** Okay, no, I think that I think that makes sense and that I can see how it'd be useful. To like for each of these data points, like what is the what is the residual and. Yeah, and incorporating more like model performance into that. Okay, and then yeah, the point with low sample is well taken. So during the development of the tool we did consider like incorporating like weighting the metrics based on like where there is data. But one concern that I had with that is I didn't want to like hide areas of overfitting. And now like for example, like I could weight the interactions measurement by like. Prioritizing areas where there's interaction in like parts of the data where there's a lot of data. But then my fear would be like that would hide cases like this where there is strong interaction. But it's not in like a very well populated area. Like it still seems. I wouldn't want to like hide someone from seeing this like once you see that it's like overfitting. It's not very interesting, but I think it's important that like you do see that like your model might be overfitting here and I wouldn't want to. Like downweight that. Just because like it's out of out of distribution. But I do also see the point that like once you identify that like yes, that's like out of distribution. Then like you might want to like going forward in your investigation like okay. I I know that now like let me see like where there's actual variation where there's data so I can I can see.

49 **Participant:** And you know so maybe just one other idea on that is I mean maybe it's almost like a tool you use in a sequence. At first you could use it to do a task like find overfitting. And then like beyond the brushing. I don't know if there could be a sense of like. Well, I know these areas are like weird and they perform weird because there're like crazy low sample on some of these important features like again like I'd have to think about if I actually think this is a good idea or

not. But I don't know if there'd ever be a reason. But kind of a use case for like kind of globally excluding some data points then some subsequent stages of evaluation. And I mean I'm guessing like I didn't really look at the code that was used to spin up this widget. But I'm guessing I could probably even if I wanted to like go back and drop data points out of whatever data I'm passing into this probably is that.

50 **Dan:** So you would have to retrain the model.

51 **Participant:** Oh, okay.

52 **Dan:** (unintelligible) like assuming that the model well. No, I guess if you did exclude those data points then it would like update the band the bounds of the PDP to not go that far so like that would be an option.

53 **Participant:** And again, like I'd have to think about it that's a good idea causing its own problems, but like. I guess just thinking about what might be a good kind of like iterative workflow for the tool is kind of interesting.

54 **Dan:** Okay. Okay, so yes, I guess in the rest of that point you mentioned. Yeah, I think it was just like sorting by the. Or having the sorting criteria weight based on the. The distributions. Was there any other point that you had there in terms of like the like running into the issue of low sample apart from that. Or other ways that you think like frequency could be made more apparent besides of the histograms.

55 **Participant:** I don't think so. And I think like to your point about making things bigger like I think also like. That's partially on me like I think if. If I made the view even bigger, viewport even bigger, so the histograms were even biggerprobably would have also been more obvious so. Yeah, lots of ways to handle that one.

56 **Dan:** Okay. Were the visualizations useful were any of them unclear.

57 **Participant:** Yeah, I don't really think of any that were unclear. I thought they were, yeah, super useful overall like with the one that I called out really easy to interpret and really. Yeah, I don't think of anything else besides the clustering that I found it unintuitive.

58 **Dan:** Okay. So for what impact did the different rankings have in your mall analysis and which rankings did you find to be the most and least useful?

59 **Participant:** I really liked the, um, I obviously the overall importance one and I think like the dissimilarity of different clusters, cause I think like that really bridges in my mind, I think the clusters can really help bridge, they're a nice stepping stone from like one-way to cluster to two-way. Um, and obviously then like the sorting then by like kind of the histogram discrepancies was great for diagnosing why some of those. I think the the line similarity one was the one that I may be um, I get the point of it, I think if I used it more, I might find more uses for it, but I think that was just the one that kept taking the most metal overhead on my part to remember what it was and why I cared about it.

60 **Dan:** Okay. Are there any other ways you think it would be helpful to rank the plots?

61 **Participant:** Not that I think of other than what we already just kind of chatted through about like, I don't know if there's like an option on the existing rankings to like. Have some sort of like volume waiting or something. Um, but. Yeah, I'd have to think more about that, but nothing else comes like. Comes to the top of my mind.

62 **Dan:** Okay.

63 **Participant:** Unless I guess going back to our other just thing about model quality. Like if there was some sort of kind of like ranking them by like correlation between like the PDP and a residual plot or something. But again, like I hesitate to say any of this stuff when I thought about it like for two minutes.

64 **Dan:** Okay. So how did analyzing subsets or clusters of instances impact your analysis? And did you find the clustering useful and the highlighting useful?

65 **Participant:** Yeah, I thought those were both really useful. And I mean, I think I'd find them more useful in a Like when I was actually trying to solve a real world problem and I had it clear why. I think at the housing example, I just kind of choked on thinking about a good like, what's something I really wanted to get into here, but I think like. I think those are definitely really well built and did a good job kind of like helping. Kind of relate different concepts and. Figure out like why things were happening. And again, I think the clustering is another place where I remember once in the housing thing. I started really getting into one cluster and then I was like, woah, this is only four data points. And so I was surprised that was even one of the ones like. That the clustering algorithm prioritized when. It is kind of like minimize variance within maximize without sort of a thing. So I think that's just something I'd want to pay more attention to in the future. But yeah, I thought those were both kind of super useful and. Well, well executed.

66 **Dan:** Great. Okay.

67 **Participant:** One thing I guess like I know the brushing wasn't linked between some of the cluster views and some of the highlighting views. And I can't and like I think usually the clusters were separated enough. It was easier to highlight in, easy enough to highlight the individual observations, but I could imagine if I had two clusters that were more like an X. That had like different trends, but we're on top of each other. It might have been harder to use the highlighting tool to investigate a cluster. So that might be something to think about. How can it be easier to like highlight a cluster for further analysis.

68 **Dan:** Okay, so for that. So I'm not sure there'd be like a good example in the datasets that we worked with, but so I guess you're saying like. What if there are cases where like you want to highlight a cluster, but like if you tried to highlight that cluster, you'd end up like highlighting other lines as well, basically.

69 **Participant:** Yeah, yeah.

70 **Dan:** Okay, yeah. Yeah, I I can see. I think that's a lot of that's at least a lot of the motivation behind like the centered ICE plot is to try to address some of that.

71 **Participant:** Oh, yeah, that's a good point.

72 **Dan:** But I think yeah, I think there could possibly still be cases where like it's difficult to single out one cluster. And then so like in this view, like the highlighting is tied to like adjusting the clusters not to like the highlighting on the one-way plots tab. But I could. I could see like there being some use in being able to like highlight a cluster from here, or maybe like. Like a button or something to like highlight this cluster in the one way plots tab and then like something like that might also help address that address that issue.

73 **Participant:** Yeah.

74 **Dan:** But did you have any other ideas in terms like how to make highlighting easier in that case.

75 **Participant:** I don't think so. I mean, I think that would be perfect or something just like some way to like. Yeah, I think something like what you said would really nicely extend kind of the existing workflow.

76 **Dan:** Okay. And then about the for the filtering capabilities. Did you find them useful and are there any additional ways you'd like to be able to filter the plots.

77 **Participant:** Yeah, no, I think that I think they were all useful. Yeah, I guess not. Yeah, we talked about I guess like options, maybe to filter data points within the plots, but I don't think of any other things of just like eliminating the plots themselves that I crave to do.

78 **Dan:** Okay, and in then terms of like filtering data points within the plots that would be about like possibly excluding the outliers, right?

79 **Participant:** Yeah.

80 **Dan:** Okay. Okay. How well did you feel the tool enabled you to analyze feature interactions?

81 **Participant:** I thought it was really good at that. As you mentioned, we called out like I think a couple of times where I just misread the axes which was on me and dumb. But I thought, especially like the differenced out view definitely just made it like easy to take in a lot of information (unintelligible).

82 **Dan:** Sorry, which view is that the what?

83 **Participant:** The like differenced out interaction. So like after like the kind of like actual, the. The kind of like orange brown one of.

84 **Dan:** Oh, okay, I see

85 **Participant:** of like after you subtract out the expected.

86 **Dan:** Right. So you like the two-way PDP that showed the interactions.

87 **Participant:** Yeah.

88 **Dan:** Okay. Okay. So moving on. Okay, so now getting to the end of the questions. What would you say are the tool's biggest weaknesses or limitations and apart from what we already discussed, are there any improvements or additional capabilities you would want PDPilot to have?

89 **Participant:** The. I think the biggest thing that comes to my mind actually is like more of a workflow question as opposed

to like feature question. And you know, especially I think (unintelligible) in our industry, we want everything to be kind of like documented. We want an audit trail to show what we looked at and thought through. So I think like any kind of like. UI based tool. I think like super low learning curve, really great to build some fast intuition. But like, I guess. When it comes to documenting this stuff, I guess I'd be curious like. How like would it be possible for somebody to just like use kind of like the individual Python functions. Like behind PDPilot to like kind of like reproduce the like plots they wanted almost in more of a like report format. Or I don't know if there's a way to kind of like. Kind of like save checkpoints as you're going along with like views of interest into some little gallery that you could like export back into a Jupyter notebook, or even just like. Click a button and have it like add a code chunk below where you're working that has the code you need to like make that view again, like I don't know how you operationalize that. But I guess just like. For us in our industry in particular. Being sure to like. Come away from it with like an artifact. I think could be like another major like accelerant.

90 **Dan:** Okay, so you, I guess to summarize that you'd like some way of either like. Ex- somehow like exporting like what you did in the tool or like exporting your findings.

91 **Participant:** Yeah.

92 **Dan:** Okay, I guess that could either be like like checkpoints like here's the stuff I looked at or like a way to like. I don't know, like download a plot that you looked at or something like that and. I guess like a way to externalize your findings. Okay. Yeah, that makes that makes sense. Okay, and then. So the last thing, this is just kind of like a catch all question in case there's anything else that. You wanted to say, but didn't apply to any other questions. So is there any other feedback about PDPilot that that you have.

93 **Participant:** I don't think of anything else right now.

94 **Dan:** Okay. So I know at one point during the study, you said you felt like you were floundering with the analysis. So is like, is there anything that PDPilot could have done better to like assist you in that?

95 **Participant:** I don't think so I think that was just more like, I don't think I'd feel that way if it was like. A model that I'd worked on. And that I knew what the like business purpose was for and therefore I knew more like how I was worried about things going wrong. But, you know, I mean, I think like when you're working with like. A toy model and it's like, I don't even know if I'm wearing more of like tax assessor hat that I'm worried about like, oh, is this going to like. Is this model discriminating? Or I don't know if I'm wearing more of a I'm Zillow and I'm trying to bid on homes hat of like, Oh, I better like. You know, like, I think. Just without a use case, it was like, there were so many things I could dig into. I think I was just like struggling to think about like. to pick which direction to go. But like, I don't think it was like an indictment of the tool at all.

96 **Dan:** Okay. And then there's one question I forgot to ask earlier. So about your like current approaches using PDP and ICE plots. Apart from you mentioned analyzing residuals were there other techniques, interpretability or explainability techniques that you commonly use?

97 **Participant:** We do a lot of like residual analysis so like, you know, fitting. Models on to the residuals of edition of current models. And kind of looking at the variable importances there to find kind of model week spots. Yeah, definitely do a lot of the. I mean, just basic variable importances in kind of PDP plots. I don't really think of anything too exciting. And again, like a lot more hypothesis driven stuff of like. How does the variable importances change in like this segment versus that segment. You know, I mean, like places where we know we have very different and heterogeneous populations, but I don't think of anything too. Cutting edge or elaborate.

98 **Dan:** All right, great. All right, so that is everything. So sorry that we went over time, but thank you so much. I really appreciate you being so generous with your time. So probably tomorrow morning, I will send the compensation. And yeah, so thank you so much. I really appreciate it. So as you know, PDPilot is open source and installable through PIP. So if you're interested in using it with your work or with your own data and model, I'm more than happy to answer any questions or help you get it set up or making changes to better support your needs. And I'm always interested in hearing about your experience findings and feedback.

99 **Participant:** Awesome. Well, thank you. This was great. And I apologize X XXXX XXX XXXXXXXX XXXX XXX XXXX XXXX XXXXXXXXXXX XXXXXX XXXXXX XXXXXXX XX XX XXXX XXXXX XXXXX

100 **Dan:** No, no worries.

101 **Participant:** XXXX XXXXX XXXXXXXX XXXXXXXXXXXXXX XXXXXX XXXXXXXXX XXXXXXX XXXXXXXXX XXXX XX XXX XXX XXXXX XXXXXX XXX XXX XXXX XX XXXX XX XXXXXXX. So again, apologies, but yeah, very cool to see. And I think like really impressive tool.

102 **Dan:** Thank you very much. All right. I hope you have a good week XXX X XXXX XXX XXXX X XXXX XXXX XXX XX XXXXXXXXX

103 **Participant:** Oh, thank you.

104 **Dan:** All right. Bye XXXX, thank you.

105 **Participant:** Bye.