

Interview

Dan: Okay. So now I just had, the last part is just an, I have some interview questions to ask you.

Participant: Okay.

Dan: And then we will be done. Okay. So I'm going to share my screen again. Okay.

Dan: So the first question is, I guess, do you use PDP and ICE plots in your work? And if so, what is your current approach to using them? So I guess the start would be like, do you use PDP and/or ICE plots in your work?

Participant: Yes, we are going to. So I'm working on the outcomes analysis for a model. So we just finished planning everything we were going to do. And so part of that includes, we're generating PDP, ICE plots, ALE plots actually too, and SHAP plots for all of the model input features.

Participant: Do I use both PDP and ICE plots? Honestly, I don't think we will use ICE plots because our sample size is so huge, we have to down sample. And ICE plots in general, I just always find very, I think we as a team find them very just hard to read and not useful and so we just opt for PDP. But then ALE can handle some of that more correlated stuff and so you look at those too.

Participant: Do we use both one way and two way plots? Primarily always look at one way. May look at two way if, for example, we have some features that. So we also look at like univariate plots with the relationship with the target and so.

Dan: What would that look like?

Participant: Hmm?

Dan: Well, what... So a univariate plot, would that just be like a scatter plot comparing like the value to the target?

Participant: Some people also call it a marginal plot, but yeah, like just looking at, so if I bin this feature into like 20 buckets, what is the average target?

Dan: I see.

Participant: Yeah. And so if I have my PDP says that the higher this variable is the higher the risk is. And if my marginal plot univariate, whatever you want to call it, says the higher this variable is the lower risk is, then that's when you might pull out the two way plot to try and understand like. is something weird going on here in how this feature is getting used in the model.

Dan: Okay, so when. So just to summarize what you said so...

Participant: So when we're trying to explain discrepancies between multiple graphs.

Dan: Got it.

Participant: Yeah. In general though, I don't think we do much (unintelligible) interactions.

Dan: So when you're trying to explain to discrepancies, how do you, do you just like compute like all the two way plots with that feature or do you, like, how do you know what second feature to put in that plot?

Participant: So we have a lot of very highly correlated features in our model and so that's usually where we would go first to try and understand. So it's like, is this feature looking this way because another highly correlated feature is getting picked up by the GBM first.

Dan: Yeah, got it. Okay, and then beyond that, do you do you look at interactions.

Participant: Not really, to be honest.

Dan: Is that because like, sorry, go ahead.

Participant: No, I was moving on to next question, but if you want to stay on this question, we can.

Dan: So for the not looking interactions, is that because like you don't anticipate the model having many interactions in that like those that like, interactions that are like worth noting or would be like uncovered through the like discrepancies between plots.

- 26 **Participant:** No, so. So we primarily use GBMs for models and I guess I don't really look at interactions for interpretability, but where I would use interactions, is we have some analysis that we do to look at what we call model weak spots. Or like. Yeah, which like helps us and the business inform like, hey, where does the business need to XXXXXXXXXXXX XXXX XX XXX XX XXX XXXXXX XXXXXXXX XXXX XXXXX XXXX XX XXXXX XXXXX XXXXXXX X XXXXXXX XXXX
- 27 **Dan:** What do you mean by XXXXXXXXXXXX XXXX XX XXX XX XXX XXXXX?
- 28 **Participant:** So our model predictions go into business analysts and then they XXXX X XXXXX XX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXX XXXX XXXXX XX XXX XX XXX XXXXXXX XXXXXXXX XX XX XXXXXXXXXXX XXX XXXX XX XXX XXXXXXX XXX XXXX XXX XXXXXXXXXXXX XX XXXXXXXXXXX XXXX XXX XXXXXXXXXXXXXXXXXXXX XX.
- 29 **Participant:** So we have some analysis we do that is like so we call it like residual analysis. And then we also have like a XXXXXXXXXXX XXXXXX. And so one is. So the residual analysis is like, okay, predict a very simple decision tree on the true target using this model and like where are pockets that some combination of features is going to improve by more value and so that we just take straight from the decision tree, like what is that, those combinations of features. So that might be what where we're looking at interactions.
- 30 **Dan:** And so the goal of that. So, sorry, so the goal of that is to determine like sub-populations with high error rates?
- 31 **Participant:** Yeah, mm hmm. Where we're either under predicting or over predicting the target, yep.
- 32 **Dan:** Okay. And then from, could you say again, how you go from like the decision trees to then like looking at interactions?
- 33 **Participant:** Yeah. So we take our model score and then we take all the model input features, like just take a really big feature set, and then fit like a basic decision tree on the target. And then like figure out how the leaves fall and those are segments that we give to the business and are like, hey, look at your program and these segments and like how this model is predicting and like, do you want to make adjustments?
- 34 **Dan:** Okay, I see. So you're basically training like a surrogate decision tree and then...
- 35 **Participant:** Yep, exactly.
- 36 **Dan:** ...from that identifying segments of a particular interest where you're like, you're over under predicting.
- 37 **Participant:** Exactly. And then we do the same thing on like, so, when we upgrade models. Also fit a decision tree where the target is the difference between my current model score and my last model score. Challenger, champion. And so that that informs the business to look at segments where as a transition models, it may, the model may be performing differently. So like just areas for them to, again, be interested in and review.
- 38 **Dan:** Got it. Interesting.
- 39 **Participant:** Yeah.
- 40 **Dan:** Okay. And then.
- 41 **Participant:** Oh, that was the third question, what tasks do you use them for. How do I determine which plots to look at?
- 42 **Dan:** And I guess before, before we move to before we move into that one, just for what tasks you use them for, could you, could you say that for one way plots as well? Did we already cover that?
- 43 **Participant:** Oh, yeah. So I mean, all of this, PDPs, ICE plots, they all come at the model validation stage. So that's where I am now. So we have a built model and then we have to do a bunch of stuff to look at. Do we trust the quality of this model? Just from an interpretability standpoint and a compliance standpoint. And that's really where PDPs come from, I guess. Does this feature's relationship with the target, is that something I understand intuitively. Based on the business. And if it's not intuitive, I better be able to explain why.
- 44 **Dan:** Right.
- 45 **Participant:** Just for compliance reasons. And then another way that we use PDPs is sensitivity analysis. And that comes into a lot of like simulation. So with XXXXXX X XXX XX XXXXXXXXXXXX XXXX XXXXXXX XXXXXXXXXXX XXXXXXXXXXX XXXXXXXXXXX And our model is like dependent on those reports to predict risk. And so like, PDPs would be helpful to say like, okay, if XXX XX XXX XXXXXXXXXXX XXXXX XXXXXXXXXXX XXXXXXXXXXX, what happens to my model prediction, can I even trust this model anymore? So those are the two main areas we would use PDPs.
- 46 **Dan:** Got it. Okay. So then how do you determine which plots to look at?

47 **Participant:** Yes. So I think we are planning to generate PDPs for all our features. But because there's 300 of them. I think we are reviewing all of them. But we would only really focus on so the most important features. So, from the GBM, get a feature importance list, focus on the top ones, and that's where most of our analysis goes.

48 **Dan:** Got it. Okay. And what would you, if any, what would you say are like the pain points in your approach to using PDP and ICE plots are?

49 **Participant:** I'm not sure. I think it's like. Like it wouldn't be something that the graph can solve, but like piecing together how all of these graphs work together to explain what is going on in the model. And then also maybe what's going on in the XXXXXXXXXX. That's more for us to focus on. I guess that's like where data scientists come in. That's like where most of the time gets spent. I don't think there's really any issues like generating all the PDPs now that we have some like internal tools to help that. So it's really like someone has to sit there and we have like PDPs, ALEs, marginal plots, and like you just sit there and compare all these charts and like make sure you understand what is going on and it's intuitive and all of that stuff. So. That's really where most of the time gets spent.

50 **Dan:** And when you say XXXXXXXXXX, what do you mean by that?

51 **Participant:** Oh, sorry. Like. The XX XXXX XXXXXXXXXXXX XX XXXX XXX XXX XXXXXXXXXXXX XX XXXX XXX XXXXX X XXXX XX XX XXXXXXXX XXX XXXXXXX XXXXXXX XXX XX XXXXXXX XXX XXXXXXXXXXXX.

52 **Dan:** Got it. Yeah. Okay. So now the rest of the questions are going to be more focused about PDPilot rather than like your general...

53 **Participant:** Okay.

54 **Dan:** ...work. So how would you say PDPilot supported or did not support your model analysis? Were there any questions you're unable to answer or tasks you were unable to perform.

55 **Participant:** Um. The only one that I can think of is like when we saw those two outliers and I wanted to like be able to pull them out. And figure out what was going on with them. But. Other than that, I think I was pretty straightforward. There wasn't any tasks I was unable to perform. I think I was just getting used to the tool. Um, I will say one thing that does bother me very slightly is on the one way plots, like if you could just increase the visual spacing between the rows, um, like my eyes just keep getting very cross-eyed in between rows.

56 **Dan:** No, that's good feedback.

57 **Participant:** Other than that, um, it was pretty straightforward. Yeah

58 **Dan:** Okay. Uh, and for so for the, uh, so this won't entirely address what you were trying to do, but, I'll pull it up just to give a demonstration. Um, so there's, so there's no way to like highlight, so you wanted to like pick out those like two data points, um, so there is no way to do that like from the two way plots tab, I didn't support any filtering there, um, but let's see, do you remember which features it was?

59 **Participant:** Oh, yeah, here. (unintelligible)

60 **Dan:** Uh, okay. Uh, above ground living area and basement finish area. Okay. Um, okay. So I think it was like these same two points. Um, so I guess what you. What you could do is that like rather than going to the two way, just go to like go to one of the features.

61 **Participant:** Yeah.

62 **Dan:** And then like select them that way.

63 **Participant:** Oh. Yeah. Yeah.

64 **Dan:** Um, and then from there you can like, from there, the like the histogram show the values of those of those two features. Um, but granted it. yeah, it's not, you would still have to like tab through to see what all the values are. Um.

65 **Participant:** Yeah, that is true. That's the, yeah, that makes sense.

66 **Dan:** Okay. So were the visualizations useful for any of them unclear.

67 **Participant:** Um, I think once I had access to the cluster plots I never had a reason really to look at the centered plots unless I was like highlighting a pocket, as you were saying. Um, and then I guess I just never use the standard plots. Um.

68 **Participant:** I love the interaction plots. Um, like the interaction ones. Sorry, the color interaction.

- 69 **Dan:** Yes. Understood.
- 70 **Participant:** Um, yeah. That was really cool. Um, and then the cluster descriptions. We didn't do too much with feature versus label. Um, so. But like that might be helpful to understanding distribution. But the cluster descriptions were like really helpful. And I assume those are sorted based on like where...what is most leading the separation in clusters.
- 71 **Dan:** Correct. Right.
- 72 **Participant:** Yeah. Like that's really cool.
- 73 **Dan:** Okay. What impact did the different rankings have on your model analysis? Uh, so which rankings did you find to be most or least useful? Uh, yes, I guess we'll start with that. So which rankings did you find to be most or least useful and what impact did they have?
- 74 **Participant:** Um, well, so I think we started with importance because in our real work models, like that is just typically where we start, based on feature importance. Um, I don't think I really used cluster difference and then used highlighted line similarity. Um, I don't think I used highlighted line similarity or cluster difference very much. I looked at highlighted histogram difference a lot. Cause I was trying to understand what a sample looked like when we spotted something interesting in the predictions. So like basically trying to profile a sample, which is something we do at work a lot too, where we look at, um, how a sample might change between models. And then, um, we have like 15 metrics that we're trying to like profile to understand like XXXXX XXXXXXXXXX XX XXXXX XXXXXXXXXX XXX XXX XXXX XXXX XXXXX XXXXXXX XXXX XXXX. Um.
- 75 **Dan:** Okay. And are there any additional ways that you would think would be helpful to rank the plots.
- 76 **Participant:** Um, maybe if you tied in like the model's feature importance, but then, like had like, give some way for the user to provide, hey, here's my feature importance list and I want the PDPs shown in this order. I don't know. But that might be a very niche use case.
- 77 **Dan:** Yeah. So my, my thought process with that is that, um. So I didn't want to rely on like model specific feature importance techniques in the tool. Because like I didn't want it to only work with GBM's, for example. Um, and then like other measures of feature importance, like based on like permutation or things like that are like pretty computationally expensive. Um, so the thought is that like rather than relying on like an additional computation, like try to calculate it based on like the ICE plots um since like...
- 78 **Participant:** Totally fair.
- 79 **Dan:** ...we're already computing (unintelligible) do that. So that was the, that was the rationale behind, behind that decision, but I can see wanting to use your own, own scores.
- 80 **Participant:** Yeah, or just like my own list of features, right? Um, so for example, like we talked about our model has 300 features. Um, like I might, if I was using this tool, like I probably wouldn't want it to calculate for all 300. It might just give you the top 50 most important features and just like plot that instead.
- 81 **Dan:** Okay, for that actually we do, we do support in that. Um, so when you calculate the plots, you do pass in a list of features that you want to calculate for. Which in this case, we just did all the features, but you, you could pass in a subset of you. If you wanted to.
- 82 **Participant:** Nice. What is the upper limit on number of rows?
- 83 **Dan:** Yeah, so the upper limit on number of rows is not super high. Um, so we much top out around like low thousands. Um, so the limitations there are, um, so one like the ICE plots get more cluttered as you as more lines, um, and then like the interactions slow down so like the brushing slows down as you add a lot more lines um so the UI gets a bit more sluggish um and then also just like slows down the computation time of the PDP and ICE plots. If you're doing it with like a lot of...
- 84 **Participant:** Yeah, I think we've noticed that because like all the data points are stored in JavaScript or something, I assume, so there's just only so much JavaScript can do before it crashes. Um, but that's like, that whole problem set of like our training sample is like 60 million rows.
- 85 **Dan:** Right.
- 86 **Participant:** And then like, I have to down sample. I think we're looking at like 6000 as our current upper limit for our PDPs and whatever plots we're drawing um and like we are fully acknowledging that like that's a severe down sample from like our (unintelligible) sample.

87 **Dan:** How are you doing the down sampling? Is it just like random or is it?

88 **Participant:** Random.

89 **Dan:** Okay. Got it.

90 **Participant:** Yep. We were thinking about doing it by segments, but. Um, that also like, you know, doubles, triples, the amount of graphs you have to read.

91 **Dan:** Right. Um, okay. So how did analyzing subsets or clusters of instances impact your analysis? Did you find a clustering and highlighting to be useful?

92 **Participant:** Yes. Um, And it really was helpful for understanding the underlying data, because we did no EDA on this data set, we just jumped straight in. So I think basically what this turned into was me using this dashboard to do my EDA, which is like not a bad way to do it.

93 **Dan:** Yeah, that is a good point. Yeah, it's somewhat like a limitation of like doing these kind of user studies is that like, it's like this is already like two hours, which is like already a lot to ask for someone to do.

94 **Participant:** Yeah.

95 **Dan:** To like really make this like a real world scenario, like you'd want them to like have a lot more experience using the tool and like seeing the data beforehand and all that stuff, which like, unfortunately, there's not not enough time to get people like fully proficient and.

96 **Participant:** Right.

97 **Dan:** With the data set.

98 **Participant:** Yeah, that's totally fair.

99 **Dan:** Yeah, it's just, yeah, it also goes with like another limitation is that like how you would analyze a model that you're entirely new to is probably like significantly different than how you would analyze a model that you work with on a daily basis.

100 **Participant:** Yeah, no, I think we did get to like my final conclusion was that, okay, this model is like really good on where there is large concentration in the data and I really do not trust this model on the edges, which is like a fair real world application.

101 **Dan:** Right. Okay. So next, so were the filtering capabilities useful to your analysis. Uh, yeah so, I guess we'll just start with that.

102 **Participant:** Yeah. Um. I think I use them in ways you didn't plan on me using them, but, I liked, um, what was it? The one way plot side. Um, I'm not sure I really used those filters too much, but on the two way plots. I really like using that. Um, because, um, that is something I want to look at is like, given this feature, I want to look at all the other interactions with it. Um, so I used that a lot. And then the two selected features, I don't think I ever used that too much, but the one plus selected feature, I really liked that one.

103 **Dan:** Okay. Are there any additional ways to filter the plots that you think would be helpful.

104 **Participant:** Oh, we did briefly talk about, well, I don't think you want to invite three way interactions in here, but that like might be something. Yeah, no, I don't think you want to, but.

105 **Dan:** So how, how would you visualize three-way interactions?

106 **Participant:** Because I know we were looking at, um, a lot of the highly correlated area features and it was just like, I want to see... Yeah, I'm not sure like how I would put another one on top of it. Yeah. Good question.

107 **Dan:** So I guess you could like.

108 **Participant:** I mean, you would use the two selected features and just have them next to each other and try to like piece together what is happening with those accounts.

109 **Dan:** Yeah, I guess you could like facet it by like the third feature.

110 **Participant:** Yeah.

111 **Dan:** Um, but I think that would probably get to be difficult to like wrap your head around.

112 **Participant:** Yeah. Yep.

113 **Dan:** Um, are there any additional ways you'd like to be able to filter the plots?

114 **Participant:** I think you had it set up very intuitively like I sometimes had to remember to click the refresh button, but, um, knowing that it was auto-sorting by like importance or like the residual essentially, um, made it very like comfortable to just like go to the tab and be like, it's automatically sorted.

115 **Dan:** So that's for the sorting, but for the, for the filtering are there any other ways to filter?

116 **Participant:** Oh, sorry, it's late.

117 **Dan:** No worries.

118 **Participant:** I can see this. Oh, wait, you said search. Use the search bar XXXX. Um, yeah, no, that makes sense. Um. No, not that I can think of honestly.

119 **Dan:** Okay. Um. All right, so you somewhay already touched upon this, but how well did the tool enable you to analyze feature interactions?

120 **Participant:** I think very well, right? Like that's primarily where we focused on.

121 **Dan:** Sorry, I accidentally set off Siri, so I wasn't able to hear that answer.

122 **Participant:** It's all good. Um. I think like the majority of our analysis was looking at interactions. And so I feel like this tool is very geared for that.

123 **Dan:** Yeah. Okay. Okay, so what would you say the biggest weaknesses or limitations with PDPilot are and are there any other improvements or capabilities that you would want the tool to have?

124 **Participant:** I mean, the alternative is just looking at PDPs on a PDF file, so, like, um, I think this is a pretty cool tool. Um, yeah. I think like the same weakness is the same weakness. PDPs have, right? Like just not being able to handle larger amounts of data. Um, so that's not something unique to your tool though. Um, any additional improvements capabilities we want to add. Um, I assume I can change the colors.

125 **Dan:** Uh, you cannot.

126 **Participant:** You cannot.

127 **Dan:** No.

128 **Participant:** Have you ever had a user that's colorblind?

129 **Dan:** Uh, so we not during the limited user studies I've done, no. Um, but I did try to pick color schemes that were as colorblind, colorblind friendly as I could.

130 **Participant:** Okay. So I'm like, you do have red and green in here, which I'm like, I know that's a pretty common color blind.

131 **Dan:** Right. But I guess that would. Yeah. So the red and green would come up if you have...

132 **Participant:** Multiple clusters. Yeah.

133 **Dan:** ...like four or five clusters...

134 **Participant:** Yeah.

135 **Dan:** Um, and like once you get to like four or five clusters, you're already like, uh, I guess like once you get to like more than like a handful of colors, you're already getting to pretty dicey territory in terms of like.

136 **Participant:** Yeah.

137 **Dan:** Finding any combination that's colorblind friendly. I guess like one, so one thing for that is that like we're not entirely relying on color alone there, um, so like within the features, like the order of the plots matches the order there. Um, so like the the rain cloud plots would fare well with that, although like with the bar charts, those would just fail because you

wouldn't be able to distinguished between the layers.

138 **Participant:** Oh, and then yes, so on the x-axis, yeah, you were like hover over it, um. Oh, did you set years to be nominal or no, no, no, they were just. Because I know your x-axis. So I saw your x-axis got butchered by year built.

139 **Dan:** Um, which one is this?

140 **Participant:** Um, so if you look at sale type, and then you look at clusters. Oh, why are you. Oh, go increase your clusters to three. Yeah, so you see your x-axis on year built. It's like, well, we have that in our charts too. Yeah.

141 **Dan:** Right, yeah, so that yes, just like in terms of like the labeling being like 1.88 K rather than 1880.

142 **Participant:** Yeah.

143 **Dan:** Yeah. Yeah, that's.

144 **Participant:** I think that happens a lot.

145 **Dan:** Right. It would just be like an annoying thing to fix to like check like is your...

146 **Participant:** Yeah.

147 **Dan:** ...number within like the specific range, then like apply this. Because then like I wouldn't want it to be the case where like. If you have like a scale that goes from like zero to like a million I wouldn't want it to be like, you know, a random like 2000 in there (unintelligible).

148 **Participant:** I think you ultimately have to have like the user tell you like.

149 **Dan:** Right.

150 **Participant:** That this is a text label or something.

151 **Dan:** Right. Yeah, that's yeah, that that's something that has annoyed me and I'd like, oh, maybe I should fix that, but then I just never have never deemed it important enough to go and dive into.

152 **Participant:** Yeah, no, it's, it's very minor. Right. Yeah.

153 **Participant:** No, I think the main, the main one for me was like getting cross eyed on the first page.

154 **Dan:** Okay. Yeah.

155 **Participant:** Because you have the expand tool on the bottom left and it like, I'm always like, is that for the one graph I'm reading or the for the one underneath I'm reading and that's just like the gap spacing.

156 **Dan:** Right. Yeah, that's, that's a good feedback. Yep. Okay. So then there's just one more question left, which is just kind of like a catch all question of like if there's any other feedback that you like to give that you didn't get the chance to be for.

157 **Participant:** So your tool is just you are you're just promoting and you would like all of us to just start using it.

158 **Dan:** Yes, this is actually a research project that I started, XX X XXXXXXXXXX XX XXXXXXXX XXX XXXX XXXXXXXX XX XXXX XX XX XXXXXXXX XXXX XXXXXXXX XXXXXXXX XXXX XXXXXXXXXXXX and now we're at like the tail end of the project, like doing the user studies and writing the paper and trying to get it, get it out the door.

159 **Participant:** That's really cool. Yeah. No, I have no other feedback.

160 **Dan:** Okay.

161 **Participant:** Yeah.

162 **Dan:** All right. So that's it. So thank you very much. I know that two hours or two hours and 20 minutes is a lot of time to donate to a research project, so I'm very grateful of you being willing to do that. Right after this, I'll send you the 40 Amazon gift card. I'll just send it to the email that you signed up with, but if you want me to send it to a different one, then just send me that email and I'll send it there. So this tool is open source and is installable through pip, so if you're interested in using for your work or with your own data and model, I'd be more than happy to answer any questions or help you get it set up. Also very happy to make any changes to the tool to support your needs and always interested in hearing about any experience, findings, feedback, etc.

163 **Participant:** Yeah. This is cool. Thank you.

164 **Dan:** So thank you very much.

165 **Participant:** I do have a question for you after you turn off the recording.

166 **Dan:** Yes, let me remember where that button is.