Interview

Dan: Um, so the first question is not, uh, about PDPilot, but it's more just about like in your work, like how you tend to use PDP and ICE plots. Um, so I guess the first question would be, like, do you use PDP and ICE plots in your work or have you at some points?

Participant: Um, I, I, no, I don't really use them. Um, as I said, on the over the, um, chat I usually use Shapley for interpretation. Um, but so, so not really PDP and ICE plots, I think my, usually my thought process there is, you know, typically we use like gradient boosting models. And because of that, there's a lot of interaction terms, especially, you know, um, without, you know, we don't really do too much like feature selection or removing of correlated variables. So there can be, um, a lot of like correlated variables, a lot of interaction. And so I guess, you know, obviously one of the restrictions of PDP and ICE plots is that they kind of just vary one feature without varying other ones so, uh, with that use SHAP, which I guess has some notion of feature importance based on, you know, feature value, given other feature values. Um, and then, you know, I guess randomization of that feature value, given other feature values and kind of subtracting them to get that kind of impact given, you know, what happens if you change that feature and given the other feature values. So yeah.

Dan: Okay. Um, so I guess to summarize that, so you tend to use SHAP, um, the thinking is that you have models where like not a lot of feature engineering. So you have a lot of correlated features and PDPs and ICE plots struggle the correlated features. So you use SHAP instead.

Participant: Yeah, well, I would say, but not, not, not a lot of feature engineering, but not a lot of maybe feature selection.

Dan: Yeah. Okay. Uh, can you talk about what you use SHAP for?

Participant: Yeah. So typically, uh, you know, Shapley for XGBoost, um, that will be, you know, looking at, you know, when you have a model. So personally in the model development phase, like trying to understand, you know, when you have a model, what features are important in that model. So ranking the feature importance and then what's the relationship between the feature and the output variable. Does that make sense? We can, from that, we can look at, you know, what features are important, where should we kind of focus a bit more over feature engineering efforts, given we see some signal in a certain data source. Uh, also are patterns as expected, um, especially in relation to null values. So I think it helps in the iteration, iterative process of model development, feature and feature engineering, do these features make sense. Okay. You know, let's iterate on them based on important ones, not important ones. Let's, let's, you know, further dig into a table or a data source based on how powerful those features are. Um, so that's the first one in model development.

Participant: And then secondly, for model monitoring, um, you know, you can have models with, uh, you know, hundreds of features in production. Um, and, you know, um, Shapley is a great tool because we can look at like drift instead of looking a feature drift, you can look at Shapley drift. And that lets you know, you know, what features have drifted significantly that may impact your model score. You know, of course, with a model with a hundred features and, you know, five of them are detected as drifting. Well, what does that mean? What does that actually mean to the model score? You know, are they important features or are they not important features? Whereas Shapley actually, you know, gives some kind of values to put to that. So, um, there are the two uses that say one is like model development, um, for understanding the model, what, what is learned. Um, do the features make sense? Um, and then a second would be for, for monitoring live models in production.

Dan: Got it. Um, so for determining what features are most important, are you calculating that based on the Shapley values or is that based on what's provided by XGBoost?

Participant: Probably both, just a sense check, but usually the Shapley values, yeah.

Dan: Okay. And then for understanding the relationship between a feature and the output, uh, can you describe how you use Shapley values for that?

Participant: Yeah. So shapley values shapley has, um, uh, like a summary plot where basically it shows, um, uh, I'm not sure. Are you familiar with the sharply summary plot?

Dan: Yes, I am.

Participant: So usually we use that. Um, some of the team members dig a bit deeper and look at some of the interaction plots. Um, but usually I just look at the Shapley summary plot to understand, I guess, uh, in, you know, relationship between certain feature and the, the score. Um, of course, you know, for decision trees, like usually, you know, the top few features in terms of feature importance, you can can think about that as what, how do these features interact with the outcome variable, but usually as you go down, then. And there's interactions with all of the features that the tree picked up later then, you know, you can't really interpret that as that because there's interaction terms present. And, um, so mostly, I would say, um, just the Shapley summary plot. Yeah.

15 **Dan:** So just so I'm clear, so like, I guess, um, so I'm assuming like not this, since that's like an instance level plot.

16 **Participant:** No. Yeah. Yeah. This next one. Yeah. This one. Yeah.

17 **Dan:** Okay. So like, what, but like, what like a plot like this? Um, I guess we can see, in general, like the distribution of Shapley values for that feature, but it doesn't really show you, like, of like how, like what the relationship for like the different values of that feature are. So I know that there are some, like other, like, like there are basically like ways, like compute, like, or like estimate PDPs from like, Shapley values. So do you, like, you look at stuff like those or?

18 **Participant:** Um, so here, so sorry, your question, your comment was about the feature, the impact of the feature on the score, is it?

19 **Dan:** Right. So the relationship between the feature and the model output.

20 **Participant:** Yeah. Yeah. Well, I mean, so here, like with the, with these plots, that's, that's color, right? From, you know, so you do get a sense of like.

21 **Dan:** Ah, Okay. Yes. I see. I see.

22 **Participant:** Yeah. Yeah.

23 **Dan:** Okay.

24 **Participant:** No, it's not, it's not as quantitative, but I think, at least personally, I'm not as interested in those details. I could just kind of want to know what's the relationship, especially where do null values fit in? Does it make sense? Are there clusters there that we should zoom in on? Yeah. Yep.

25 **Dan:** Okay. Got it. Um,

26 **Participant:** But I would, I would definitely consider using this tool as a compliment to that, because I think you, I would get insights from this tool, uh, you know, using it even just like for a trained model, like doing 15 minutes of analysis with this, I think it would also complement that.

27 **Dan:** Great. Um, so in your current approach with using SHAP values, do you, are there any particular pain points?

28 **Participant:** Uh, well, I mean, just in general, you know, model interpretability and, and trying to know what a model does, uh, is not, it's not straightforward. You know, most methods rely on some kind of heuristics or assumptions or whatever it is, you know, for Shapley plots, they can actually be quite dangerous for interpretation, because stakeholders can say, Oh, well, this feature here, correlates, you know, these, these, like customers with these attributes have whatever higher churn probability because this is what the SHAP plot looks like, which is a totally wrong interpretation.

29 **Dan:** Right.

30 **Participant:** Um, so, pain points. Um, I wouldn't say pain points as such. I just, I would say more limitations, like, you know, I think these models are very complex. Um, you know, like, you know, gradient boosted machine, is like an ensemble of whatever, hundreds of trees that are ensembled in a kind of a nonlinear way compared to like random forest classifiers. So it's like, all right, trying to understand what the model does is, uh, is kind of tricky, but I would say like, I think the use cases are understanding how you can have more performant model by understanding relationships between variables and, uh, outcome variable, like what's your model not picking up on? Are you tuning it in correctly? Is it overfitting those kinds of things? Um.

31 **Dan:** I see.

32 **Participant:** So I think it can be useful for understanding like how to tweak the model parameters, maybe other feature engineering. Um, yeah, just, yeah, from you and your experience and what you've talked people, like why, what, what are problems are people typically trying to solve with like feature interpretation and.

33 **Dan:** Uh, so I think one of the problems that we're trying to address with PDPilot is like, uh, especially with models (unintelligible) like that have many features, people can kind of be like drowning in the number of plots that they can possibly generate. So like the idea is to help people have like a way to like more efficiently find the plots that are interesting and worth looking at.

34 **Participant:** Yeah. Yeah. And, but what, what are people trying to do with that information? Are they trying to improve the model or something else? Is that usually the thing to try to improve, the model?

35 **Dan:** Uh, yeah. So it would be a mix of like, like during model development, but also like, during like model validation,

um, to like verify, especially like, uh, if there needs to be like compliance checks with this model or like other, um, other kind of like audits of the model to make sure that like it's not doing anything too strange.

36 **Participant:** Yeah. Yeah. Got it.

37 **Dan:** Okay. So next I have some questions about PDPilot in particular. So the first one just starts off pretty generally. Um, how did PDPilot support or not support your model analysis? Uh, were there any questions that you're unable to perform or any tasks, or sorry, any questions you weren't able to answer or any tasks that you weren't able to perform?

38 **Participant:** No, it's pretty, it was definitely it was definitely quite good. Um, I think obviously I had some comments about maybe the ranking systems or maybe weighting by number of data points and so on. You know, yeah, I think maybe some of them might have been skewed by those outliers. So maybe that kind of, I would say maybe not that I was unable to answer, but maybe I think the plots led me down, maybe a bit of a rabbit hole where there's just a few outliers. Obviously that's, you know, important to like, hey, your model can kind of do some, you can iterate by maybe looking at outlier detection, but, um.

39 **Dan:** Right. So that's something that we considered in that, um, so I considered adding functionality that would weight the, the calculations by the number of data points in that area. Um, but the fear is that that could make people like miss outliers. Um, and now they, in this case, like with the model that we're exploring, it does seem important to know that your model's overfitting on three data points. And I wouldn't want to like hide or downplay that overfitting.

40 **Participant:** Yeah. Yeah. Yeah. Yeah.

41 **Dan:** But I can understand that like once you've identified that overfitting, um...

42 **Participant:** So you're saying like, this tool should show you is important information that maybe your model is overfit, and then you should retrain and then in the next iteration without that overfitting (unintelligible) you know, the charts (unintelligible) more regular because there wouldn't be those outliers.

43 **Dan:** Right. Right.

44 **Participant:** The model has fit, some crazy, you know, yeah. Yeah. Yeah. Okay.

45 **Dan:** Yeah. That, yeah. That's the idea. But I, I can see the, the usefulness in like, being able to like downweight by. Uh, yeah. Okay.

46 **Participant:** Yeah. Cool. Got it. Um, a task unable to perform. Um, I mean, I think wanting those particularly, and this is more probably of a EDA question in general, especially on the first data set, the bike sharing data set is more, you know, you've a lot of correlated variables that kind of, um, make interpretation a bit harder, like there, and it was like what time since 2011, and it seems to be (unintelligible) like other things. So, you know, I think like, trying to understand those clusters of correlated variables, maybe which ones are actually kind of, obviously this is hard, but maybe it's more, you know, trying to understand which ones are actually important to look at, whereas which one's actually correlating with the important one to look at, that's impacting, you know, but that's a harder, that's, that's like, maybe even a causality question or something, but,

47 **Participant:** You know, actually, I have some experience with, um, like interpretable machine learning methods, and, uh, XXXXXXX X XXXX XX XX X XXXXXXX XX XXXX XXXXX XXXX XXXX XXX XXX XX XXX XX XXXXXX XXXXX XXX XX XXX XXXXXX XXXXX X XXXXX XXX XX XXXX XXX XXXX X XXXXXXX XXXXXXXX XXXXXX XX XXXX XX XXX XXXXXXXXXXXXX XXXXX XXX XXXXXXXXXX XXXX XX XXXX XXXXX XXXXXXXX XXXXX XXX XXXX, I guess some stuff like that, but it's, uh, it's a hard, you know, always one of the problems I faced with this as well was like, you have so many features and they're correlated and how do you kind of, you know, pick one of them to represent the correlated class or something, or I don't, I'm not too sure, but it's, that's kind of one thing that's tricky as well, with, with, with feature analysis.

48 **Dan:** Got it. Okay. Um, so for the visualizations, uh, did you find them useful and did you find any of them unclear?

49 **Participant:** Um, no, they were useful. Yes, they were useful. Uh, unclear? No, well I mean, you walked me through it very comprehensively, so it wasn't unclear, no.

50 **Dan:** Okay.

51 **Participant:** It could be a better question, it could be good to ask that to somebody who just, you know, uses the package and reads the documentation, but, you know, yeah, you, you asked me, so yeah.

52 **Dan:** Okay. Um, what impact do the different rankings have on your analysis, uh, and which rankings did you find to be the most really useful?

53 **Participant:** Um, well, the histogram ranking one is very useful. Yeah. Uh, and you know, the importance one obviously is useful. I think I would still complement the importance one or sense check that without other, like, like a SHAP feature importance or like, uh, logistic regression coefficient analysis or whatever to sense check it. Um, but yeah, I would say, yeah, the ranking least useful. I think the clustering one, I didn't really, maybe, maybe I don't fully understand it, but the clustered lines, how they kind of clustered together.

54 **Dan:** Was that with the highlighting one or the non highlighting one?

55 **Participant:** The highlighting one, I guess, yeah.

56 **Dan:** Okay. Uh, and I also noticed that you didn't really use the, the cluster distance one.

57 **Participant:** Yeah, could you remind me what that one did, maybe (unintelligible), I think it's just because we went through so much, so much information. But can you remind me what that one did?

58 **Dan:** Uh, yeah. So for that one, it's about, um, um, ranking the features based on how far apart the clusters are.

59 **Participant:** Oh, yeah. Oh, I think that would have been useful. That would have been useful. Yeah. Let me see. Um, yeah. That would have been useful, I guess. I just, um, you got second floor area, year remodel. Yeah. Very useful. Yeah.

60 **Dan:** Cause yeah, there was a point where you were doing this type of analysis in like, trying to find like different clusters, but didn't rank by that. So I wasn't sure if there was a particular.

61 **Participant:** No.

62 **Dan:** Okay. Um, and so, so I also noticed that, at points you seemed to be preferring the XGBoost rankings over the, uh, the tool's feature importance rankings.

63 **Participant:** Yeah.

64 **Dan:** Um, so I guess I was wondering if there is like a particular reason for, for that. Because like, I know like at the beginning, you wanted to do like a sanity check to see if they were similar, but even after seeing they were similar, you seemed to like regularly scroll down to the XGBoost list to see which features to look at and then go back up to the tool.

65 **Participant:** Yeah. Well, I mean, I guess it's the ground truth, you know, it's like, that's the how the model is learned. Um, so I think that's what I would kind of consider as the ground truth, as such, um, you know, that's literally like, you know, you know, when the tree produces the split and you get a gain in the performance it's like literally that much. It kind of tells overall like which features are (unintelligible).

66 **Dan:** Right.

67 **Participant:** You know, I think maybe because again, maybe the, the, the I see lines are weighted, are kind of aren't weighted by the data distribution. I think maybe that's one reason, maybe I would kind of distrust those rankings and just keep referencing the other ones, but, um, yeah, I guess I was always kind of sense checking with the XGBoost ones to make sure I was looking at the most important features. Yeah.

68 **Dan:** Okay, got it. So do you think would it, would you find it useful if there was a way to provide like your own feature importance rankings to the tool?

69 **Participant:** Um, yeah, I think that would be useful. I think that would be quite useful. Um, now, in this scenario. Yeah, but maybe, maybe, maybe also just that's another level of complexity, like in this, yeah, maybe like, I'm just thinking out here, you know, for me, it wasn't hard to reference that like feature importance dataframe. And then your tool makes it very easy to find features and select. So, so maybe, but I wouldn't say it's essential or anything. Yeah.

70 **Dan:** Right. I guess it would just take, take out the need to like go back and forth between the two.

71 **Participant:** Somewhat. Yeah. Yeah. Yeah.

72 **Dan:** Um, okay. Are there any other ways you think it would be helpful to rank the plots? I guess apart from the, the outlier sensitive one and

73 **Participant:** No, I think the other thing that was missing actually another thing about in the analysis was, you know, when I was zooming in on those that weird area plot where everything jumped, I'd like, I'd like to be able to highlight the histograms rather than just the curves, you know, when you, you know, your highlighting tool I guess highlights instances.

74 **Dan:** Right. So...

75 **Participant:** I guess I also wanted to highlight instances, but based on their feature value.

76 **Dan:** Yeah. So for that, you could have, um, in this histogram, sorry, in this scatter plot, you can brush the scatter plot to highlight those points. Um, but you can only do that from the detail plot tab. You can't do that by the histogram.

77 **Participant:** So if you, if you go back to the standard. Okay. Got it. And then can you do sorts by highlighted histogram difference. Got it. Okay. That's useful. Yeah. Yeah. Okay. Yeah. So that's something I just guess I didn't know yet. Yeah. Yeah.

78 **Dan:** Okay. Um, so how did analyzing subsets or clusters of instances impact your analysis? And did you find the clustering useful? Uh, and did you find the highlighting useful?

79 **Participant:** Yeah. I think they were both useful. I think the highlighting was probably more useful, but the clustering is also quite useful. Yeah.

80 **Dan:** So is there a reason why you preferred the highlighting over the clustering? So I guess in particular, at the beginning, um, I'm not saying there's like any right or wrong way to do this. Uh, so I think the approach that you took made a lot of sense. Um, but for like, for this one, for example, so like you like brushed that and then sorted. And then you also like brushed that and then sorted, um, which like definitely works. I guess another way would be to like look at the clusters and then view it through this, uh, since like that kind of like already has that selected and that selected to, to analyze. Uh, so just like an alternative way of doing the same thing, but I think you would arrive at the, at the same conclusion.

81 **Participant:** Cause you can manually do the clusters as well, can't you?

82 **Dan:** Yes.

83 **Participant:** Yeah. So I think that way would have been better, yeah to do it that way.

84 **Dan:** Okay. Uh, so was there a particular reason why you defaulted to the, to the highlighting? Was that just because of, uh, like what came to mind first or?

85 **Participant:** Yeah. That's what came to mind first. Yeah. Yeah.

86 **Dan:** Okay. Um, so were the filtering capabilities useful for your analysis? And are there any additional ways that you think would be useful to filter the plots?

87 **Participant:** Maybe you could have, like, generally everything is maybe like, no, actually, no, I don't think so. Yeah. Uh, no, I think, I don't know. I think it was, I think it was, I think they were, they were very useful. Yeah. I can't think of other ways.

88 **Dan:** Okay. Um, so how well did the tool enable you to analyze feature interactions?

89 **Participant:** Um, yeah, good. It was easy. Um, I think it is cool the way everything is pre-computed and then you just calculate them. So, um, it was quick versus like say a SHAP interaction plot where you have to like compute everything on the go. So I think that like pre-computing everything and then being able to dive in is pretty neat.

90 **Dan:** Right. Uh, yeah. So like I guess the downside of PDP and ICE pots is that it requires a lot of queries to the model. So they can be like pretty slow to compute. Um, so I guess like the motivation for pre-computing it was mostly to cut down on like, like I wouldn't want us to be sitting here for five minutes waiting for.

91 **Participant:** Oh, no, on a general, like, you know, when you're looking at your examining things, like if you have to wait every time you compute something (unintelligible) 20 seconds, then that's slows the rest, so the factor of pre-compute is definitely a nice design element.

92 **Dan:** I see got it

93 **Participant:** As I said, I think like the raw, when it came to the 2D plots and 1D plots, like showing the the raw, y-values is useful because I could see what outliers there was, but I think also showing like the, um, maybe distribution side of that too, like, uh, could be useful too, because I'd like to be able to compare the prediction plot to the average outcome variable plot to see are there certain regions where the model is not as accurate. Does that make sense?

94 **Dan:** Okay. Uh, so for that, um, (unintelligible) go to a two-way PDP, um, maybe that's not the best example. Uh, so like here, you're saying that like it would look like this, but it'd be computed based on the ground truth labels as opposed to, uh, it's like it'd still be a heatmap, but then like in each bin, I would take like the average ground truth label.

95 **Participant:** Yeah. Yeah. Yeah. Yeah. Precisely. That would be a nice option to have. And, and, and where there's no

data points, uh, having them as missing as well, or just white space as well, I think is useful.

96 **Dan:** Right. So is there a particular reason why you prefer that over showing the instances?

97 **Participant:** Um, it's just it can be maybe noisy. Like if you go back to, uh, like it does it well where the data points are there are lots of them, but you know, where the purple orange, where the cloud is less dense, it's kind of harder to make up the pattern, I think. So, you know, I'd love to basically see like a prediction plot and the other one and see is the pattern the same. And if not, maybe it's (unintelligible) to a certain area of the feature space where the model is not capturing the the risk the same.

98 **Dan:** Okay. Um, Okay. So these, entering into the last questions. Um, what would you say are the tool's big, this weaknesses or limitations? And are there any, uh, improvements or additional capabilities that you would want the tool to have?

99 **Participant:** I suppose like back to my earlier point about, you know, partial dependence plots and ICE plots of like, it could be useful maybe to have some metrics of saying like how, I don't know if these metrics exist, but like how much to, you know, are these curves, the ICE curves, trustworthy or are the feature interactions too strong that maybe make them less trustworthy, something like that. Maybe to know, um, maybe that's useful. You know, I think you going into model interpretability, you always kind of know nothing's perfect, but you can find some information that you can validate, maybe through the raw data as well afterwards, but so maybe but maybe not. Um, no, I think everything I've said in terms of improvements is just small things, but I think, um, yeah, it's nice. It's got a package that you can easily install and easily use. So, yeah, seems pretty cool to me.

100 **Dan:** All right. Great. So now this last one is just kind of like a catch all question in case there's anything else you wanted to say that you didn't have a place to say in previous questions.

101 **Participant:** Um, no, I think, yeah, as I said, I think it would be useful to consider null values because especially with like, XGBoost and stuff like that, like we don't impute null value. Um, so I think that could be a gap for models where there are, you know, null values that are important. Um, but that's the only other thing. I think, yeah, that's just re-emphasizing. Yeah.

102 **Dan:** All right. So thank you. Um, I really appreciate you donating your time. Um, for this, I know that this is a lot of time to spend for my research project. I'm very grateful that you're willing to do that. Um, so right after this, I'll send you your compensation. XXX XX XXXX XX XXXXXXXXX XXX XXXXX XX XXX XXX XXXXX XXX XXXX X XXXXXXX XXXX XXXXXX XXXXXXX XX XXX XXX XXXXXXXX XXXX XXXX XXX XXXX XX XXXXXXXX XX XXXXX XXX XX X XXXXX XXX XXXXXXXX XX XXXXXX XXXXX XXXX XXXX XXX XXX XXXX XXXX X XXXXX XXX XXXXXXXX XXXX XXX XXXX XXXXX XXX XXXX X XXXXX XX XXXXXXXXXXX XXX XXX XXXXX XXX XXX XX XXX XXXX XXXX XXXXXXXXXXX XXX XXXX XXXXXX XXXXXXXX XX XX XXXt

103 **Participant:** XXXXXX XXXXX XXXXXX XXXXXX XXXXX XXXXXX XXXXX XXX XX XXX XXXX XX XXX XXXXX XXXX XXXX XXX XX XXXXXX XX XXXXXX XX XXXXX XXXXXXX XXXXXXX XX XXXXXXX XXXX

104 **Dan:** Yes, absolutely.

105 **Participant:** Yeah, that'll be great.

106 **Dan:** Um, yeah. And then the last point, so as you saw, PDPilot, it's open source and it's soluble through PIP. Um, so if you're interested in using it for your work or with your own data and model, uh, I'd be more than happy to answer any questions or help you get set up or make any judgments that PD pilots support your needs. And I'm always interested in hearing, uh, about your experience and any findings and feedback. Uh, yeah. So thank you so much. I really appreciate you participating.

107 **Participant:** Yeah, that's for the time. I learned a lot. So thanks. Thanks. I appreciate it.

108 **Dan:** All right. Thank you, XXXX. Hope you have a good day.

109 **Participant:** You too, Daniel and best of luck with it. Thanks a lot. Great work. Thank you.

110 **Dan:** Bye.