

Levi: Hi guys, thanks for joining us. This is the hands-on machine learning broadcast. This is Taylor Larson.

Taylor: Yup.

Levi: I'm Levi Thatcher. Taylor's joining us for the first time. He's on the data science team.

And Taylor, how are you doing?

Taylor: Yeah, I'm doing good.

Levi: Good, good.

Taylor: Thanks for having me on today. I'm just going to sit in for Mike today. I'm not a Mike 2.0 but—

Levi: Yeah, yeah. No,—

Taylor: --glad to be sitting in.

Levi: We're glad to have you here. So what's going on?

Taylor: Not much.

Today, we've got some exciting things on tab. But, before we dive into all of that, I just want to mention that if you'd like to contribute to the chat window be sure to log into YouTube. We're going to maybe be sharing a couple of things on the screen, so adjust your video resolution if you need to and don't forget to subscribe to the channel as well as our blog so. That way, can follow on and get some good notifications as we go so.

Levi: Yeah. On the YouTube though, you can turn up the high resolution. I think that kind of helps with what we show here.

Taylor: Yup.

Levi: So yeah, check that out. And then, in terms of what's on the docket, mailbag first?

Taylor: Yup, mailbag. Then, we're going to do some machine learning in the news like you guys normally do. And then we are going to chat about some healthcare use cases for healthcare.ai. And then we're going to open up the chat and just try and get

any questions answered or kind of have some feedback going on, so it should be good.

Levi: Yeah, yeah. That's awesome. And we want to say that this is nothing without you all. We want to really make this a community and learn from you and hear like what problems you're facing in your data work. So, please, log into the chat. Let us know or you can email us. There's a contact page on Healthcare.ai, so check that out.

And let's get start with the mailbag. Do you have those questions up or should I pull them up over here?

Taylor: Yeah, I've got them.

Levi: Cool. Let's see what we've got this week?

Taylor: Yeah. So, the first one we had, that one came from Samir - emailed to healthcare.ai. He was working on a project in a healthcare system and has some real-time data - things like temperature, heart rate, oxygen level, and wanted to know what type of model would be a good use for that and kind of how to set it up so.

Levi: Oh, that's a really interesting use case.

Taylor: Yup.

Levi: So, of course, in a health system when you're dealing with an EMR or health metrics in general your measurements - the things that you're using to train a model, these could be hourly, daily, or even weekly. And so, how do you kind of like-- how do you start with that? How do you kind of get them all together? Like what—

Taylor: Yeah, it gets kind of complicated but it's exciting data to have - potentially, some really useful data. But yeah, if you've got-- different people in the hospital-- one person might be getting their temperature checked every couple of hours, so they've got many, many records of their temperature. Where then, someone might just have their temperature checked a couple of times during their hospital stay. And so, you've got to aggregate that data up to some sort of grain, kind of, for the machine learning algorithm to be able to handle it handle it in a standard way. And so, some ways that we like to do that include things like maybe the earliest temperature reading, the latest, the highest, the lowest.

Levi: Oh yeah, all sorts of things you can do with that – the mean.

Taylor: Yes.

Yup, exactly. Just get it to some sort of, you know— so that your columns don't have to be infinitely wide because you don't know if the person's going to have 4 readings or 24 readings.

Levi: That's a good point.

Taylor: Yup.

Levi: So, I had never really heard of that concept “grain” before I came to healthcare.

So, Taylor comes from Colorado. And when you were Colorado Medicaid, was that concept used?

Taylor: Yeah. Yeah. You know, you kind of have some examples of grains are like the patient level. Maybe you're also talking about the visit level. So, one patient can have multiple visits. Then one visit could have multiple days, so you might have a day grain. And so, that's kind of the concept of grain. And thinking about that you want all of your rows for machine learning to be at the same grain so that you can feed it in accordingly.

Levi: Okay, that's awesome.

Taylor: Yeah.

Levi: Yeah. So basically what does a row represent? Is that a an hour, a day, or an encounter – basically? That's great.

Taylor: Yeah. And I think that those other data fields that Samir mentioned - heart rate and oxygen level, those would kind of follow the same format. And you just sort of really want to think about and probably - at least, in my case, I'd need to chat with someone that knows much more about the clinical setting than I do and ask for some advice on what would be the most clinically relevant handful of fields that you can kind of start with? And then test those in your model to see how those help.

Levi: Yeah, it's definitely helpful when you can kind of pair the analyst and the data analysis with the clinical expert. That's the way we try to do things here at Health Catalyst.

Any other questions come across?

Taylor: Yeah, yeah. We had a couple others. Another good one that we'd actually seen on a couple of occasions was "How does forecasting differ from machine-learning? And that can the same algorithms be used?"

Levi: Yeah, that's a great question. So, we've heard that a lot and forecasting sounds pretty similar to prediction and machine-learning. You're looking out into the future. And how we think about it is, with machine learning, we often are predicting kind of a "yes – no". "Will this patient get [inaudible 00:05:47] kind of on a personal level. Whereas, forecasting seems to roll up to, "Okay, well how many beds will be utilized in this department, this day?" Almost more of like a summary type calculation. And perhaps it'll bring in seasonal effects a lot more than you would with typical machine learning. So, you know, substance doesn't care what day of the week it is. But if you're looking to predict, "Okay, well how busy is this department?" You'll definitely want to bring in things like month of year, day of the week - things like that.

Taylor: Yeah. Yeah. And it seems to be especially a popular use case for operations for finance to know "How many beds are going to be used? How many people are going to come in with this specific condition?" So that they can, kind of, do a bit more resource planning. Or "what is our cost for this department going to be for a certain amount of time?" And yeah, it's definitely kind of a little bit higher level, kind of the aggregation of those kinds of grains that I mentioned earlier.

Levi: Yeah, so you can use the same algorithms because, essentially, just rolling up to different grain.

Taylor: Yup, yup. I think, as long as you set your outcome variable up correctly, to be out in the timeframe that you actually need –how likely is this person to come back to the hospital within 30 days, you know?" Now, you know when they've been discharged. You know what their likelihood is. And you kind of create that kind of a moving 30-day window or something like that. And at least to me that's—

Levi: Yeah, yeah.

Taylor: The use of forecasting—

Levi: No, exactly.

Taylor: With the algorithms that already exist.

Levi: Yeah, so it seems really valuable in more of the operational and financial sense—

Taylor: Especially—

Levi: A lot more than in the clinical sense.

Taylor: Yeah, yeah. A lot of the use cases, which we'll kind of get to a little bit later in the show, end up really wanting to focus on the patient which I think is really great but then there's other parts of the health system that have to be accounting for the whole system and planning accordingly.

Levi: Yeah, yeah. Fantastic questions. Keep them coming, guys. So we really appreciate you reaching out and love digging into what you're up to out there.

Taylor: Yeah.

Levi: Anything else going on?

Taylor: Yeah, we have one more about defining and the training versus the testing data for the— it's something like a 30-day re-admission.

Levi: Yeah, that's a great question. So whenever you're doing prediction you'll have a training set and a test set. We've gone through this a little bit. But on a high level, the algorithm needs to learn from past folks - what were their attributes? Did they have a good outcome or not? And then you're predicting on a different set of folks. And the idea with this training set is those are the people that you're learning from. And the test set are those people that you're giving a prediction for. So each day, if you're predicting central line infection, your test set are those people that have a central line in them. And the training set are all past folks that have had a central line in the past.

Taylor: Yup.

Levi: And so, when you've done this for a 30-day re-admission. It's fairly straightforward, just kind of with a time window there.

Taylor: Yeah. There's a couple of time windows that come into effect with the 30-day re-admission, kind of depending on the use case where you've got patients that are still in the hospital. You'd like to know how likely they are to be re-admitted after they've been discharged. And then you also have patients that have been discharged but they're still within a 30-day window. They have not yet been re-admitted to the hospital but they're outside of the hospital so maybe there are some other interventions that come into play.

Basically, the test window for that would be anyone who has not been discharged more than 30 days and who has not already been re-admitted. And those would be the patients that would be your test window. You would predict on them. And everyone else that either has already been re-admitted or is outside of that 30 days - those would be your training set.

Levi: Yeah. That's awesome. That's a really practical question. And you can kind of apply those principles to many different use cases.

That's the nice thing with machine learning. You really don't need a PhD about it. You can kind of learn how these concepts work and then say, "Okay, well, with this particular business question, let's map what we're doing in the past, what we're doing today." And a lot of these things are very symmetric across use cases.

Taylor: Yup. And if you definitely just kind of look closely at your use case, rely on your personal experience with the data and just make sure that it makes logical and practical sense. And that's kind of the best way to start. And then you can work on refining that definition a little bit if you need to. But relying on your own expertise is a great place to start.

Levi: Yeah, yeah. Thanks so much guys for reaching out. We love hearing from you.

And again, whether it's in the chat window or in the contact page on our website, criticism, dress - fashion sense recommendations. I could use that—

Taylor: Yeah. I don't know if this shirt was too busy.

Levi: -- so that's nice, that's nice.

Taylor: But I threw it on anyway.

Levi: [inaudible 00:10:33]. I like it.

Taylor: Yup.

Levi: We're happy to take anything you've got out there.

And so, that's the mailbag for today. So where should we go next?

Taylor: Yeah. And now we're onto machine learning and the news. And I'd heard something in the news that Google had acquired Kaggle, is that right?

Levi: Kaggle. Yeah, yeah.

Taylor: What is Kaggle?

Levi: Kaggle. So, not Kegel. You know, there's some mispronunciation out there.

Taylor: Yeah. I'd heard it on a previous show that I was watching so I didn't want to go down that path again, so.

Levi: Yeah. Yeah. I believe it's Kaggle.

Taylor: Yeah.

Levi: Okay. And so Kaggle – we'll show here on the screen. Kaggle is this amazing machine-learning website where you can get your hands dirty with data and you can find data sets that you can download and play with. And what you essentially do is there's a little competition where you compete against other folks in the field. These could be researchers, or people in their basement, or whoever. And you're all trying to predict something that's in this data set.

And so, if we go to this competition's tab really quick, we can kind of show you how that works. There's always a bunch of competitions open at one time. And as we scroll down, you'll notice, okay, well, something to do with fisheries that you're predicting or YouTube video predictions. And you'll notice there's quite a bit of prize money involved in these competitions but there's also ones that are just here for practice.

That's actually how I got my start in machine learning. And I finished my PhD, went into an analyst role and was learning SQL. And I was eager to learn about machine learning and I was getting kind of this buzz around it. And I was like, "Wait. Well, it sounds like statistics."

And here at Kaggle, you can practice the things you're hanging on the show. You can try different algorithms. And so, check it out. It's fantastic.

And Google bought it. You know, they're buying quite a few type companies. So they bought DeepMind a couple of years back which is a British AI firm that actually taught a computer how to play Go. Have you ever heard of Go or played Go?

Taylor: I don't think so.

Levi: I had never heard about it either until I read about the work they were doing in Britain but Go is this ancient Chinese game. It's fairly simple but like has an

incredible number of perturbations about it. And so, it's really quite a feat that the computer learned how to play Go well. But getting back—

Taylor: Is that the one where it got more aggressive over time or something like that? [inaudible 00:12:54]

Levi: Oh, oh. Yeah, yeah.

Taylor: [inaudible 00:12:56] different game but yeah.

Levi: That's not something related to that racist bot out there, right? There was a racist bot I heard that--

Taylor: Yeah. No, I don't know about that one.

Levi: Yeah, that was the aggressive thing that like connected.

Taylor: Okay.

Yeah, with that Kaggle and Google, do you have any other thoughts on maybe why they did the acquisition? I'd read a little bit that they're interested in getting access to more data scientists and—

Levi: Yeah.

Taylor: Kind of getting a foot in the door on that growing field so [inaudible 00:13:23]

Levi: Yeah, yeah. And Kaggle- that's a good question. So Kaggle's kind of the center of public data science, you can think of it as. They have data sets. They have competitions. They have kernels which might sound a little bit abstract but if you go to the kernels tab, it basically gives you a bunch of different sets of code. They'll let you see how to do analysis on this or that kind of data.

Taylor: Oh, cool.

Levi: Yeah, in Python, and R and other languages, so definitely check it out.

Taylor: Yup.

Levi: So that's what's in the news this week. So the heart of today's show is going to be around some of the work that Taylor and I'm doing and others on the data science team where we've been actually going across a lot of different use cases in healthcare and saying, "Well, how does the machine learning fit into this or that?"

And what type of machine learning goes in here? And how do you actually make an impact on these outcomes improvement projects?

So, Taylor, do you want to run down a couple of these different use cases that you find that are like different categories of them?

Taylor: Yeah, yeah. I felt like at least in the type of work that we do where most of our work is around kind of an outcomes improvement methodology. And so, all of the projects kind of tie out to a few different categories that I'll go over but they all have to, of course, drive some sort of outcome either some sort of operational efficiency, financial efficiency and then definitely the one that we focus on a lot is clinical outcomes improvement. So those are kind of the three categories – the operations and performance management, finance, and clinical. And then I think that we can break clinical down even a little bit further into acute and chronic. So if you want, I can just chat about a couple of use cases—

Levi: Yeah, yeah. Let's—

Taylor: And we could even dive into one deeper, if you like?

Levi: Yeah. No, that's awesome.

Taylor: Cool.

Levi: It's kind of nice to get the high-level categories and maybe kind of explore some of the sub-categories therein?

Taylor: Yeah. For sure, yeah.

So an example of an operations and performance management one might be kind of at the practice level - the outpatient practice level around appointment no shows where there's some lost efficiency. And some folks that aren't getting the proper follow-up care, if they're not showing up to their primary care appointments. And so, that's a great spot for predicting whether or not patients will show up to their appointment. And then either use that to increase the intervention of maybe calling them, or following up, or helping work around their schedule or transportation needs, or improving your scheduling and efficiency as far as double booking certain slots. And there's a lot of different ways to use - what seems like one prediction can have many different use cases within that application.

So, another area is finance. Like I mentioned, propensity to pay - how likely someone is to pay for their medical care. And this is an area that can also help with find out what resources maybe a patient needs the hospital system to help with, whether it's maybe timing of the payment, or some other sorts of arrangements, or just maybe a reminder that the bill had gone out to a bad address and you can follow up so.

Levi: Yes, a lot of different benefits right there. It's not just about the hospital deciding who to deny care to. It's more, I would say, mutually beneficial as with a lot of these imbalances that are maybe— say it's \$6000 and the person can't pay that. But maybe if they do some sort of balance modification the person could pay \$3000 and then both parties benefit.

Taylor: Exactly. Not everything is going to bad debt. You could have just have a lot better outcomes for the hospital system and for the patient which is kind of the ultimate goal.

And then on the clinical side, there's some really interesting use cases specifically around acute care. Hospital-acquired conditions is an interesting one. Things like CLABSI which I know we've touched on before. Hospital-acquired pneumonia. Sepsis - that's an interesting use case.

And that could be anything from predicting whether or not the patient will get that condition or helping identify that they have that condition and maybe were misdiagnosed with something else. And so, helping kind of hone in on a bit of a bit of a registry which is really nice. And then, of course, you can tie some of our most common outcome variables to those cohorts as well, so predicting length of stay for an acute condition, predicting re-admissions after they've had that condition.

Levi: So, it's kind of like this matrix, you can kind of like apply a lot of different scenarios for focusing on heart failure, for example.

Taylor: Yup.

Yeah, so you can go from predicting the condition itself. A similar thing would work for chronic conditions or predicting some sort of outcome following a procedure, a condition or whatever have you.

And like I said, same with on the on the chronic side, you could use predictive analytics to help with registry prediction or any of those other outcomes. And

we've definitely got into a lot of those before – COPD, heart failure, diabetes - things like that.

Levi: Yeah, that's fantastic. So, it really boils down to what is most beneficial to your business goals. Of course, you're looking to increase the efficiency of patient care and the hospitals love saving money, helping the bottom line. And so, we kind of let that drive the discussion here. We're the ones saying, “Okay, well, it's great to create models but if they're not doing things that are helpful to the business then they're really not really helpful at all. I mean, it's just kind of an exercise— it's more of an academic exercise.

Taylor: Yup.

Levi: So, we're practical focused and love saying, “Okay. Well, how can we help this unit be better?” And then kind of say, “Okay, well, here are the scenarios where machine learning could come in and help - whether it's length of stay, like he was saying, or re-admissions prediction, or even the disease itself.

So, CLABSI's a model that we found to be quite accurate. And it's interesting because it doesn't bring in a lot of those longer term variables. CLABSI is a central line infection. It doesn't really matter so much whether you are of a certain weight or have high blood pressure. So what's kind of the distinction there is that's more of like in-hospital.

Taylor: Yeah, so I think if you're kind of looking at a use case where you're trying to predict a specific condition then as your consulting with the clinician or something, maybe for your CLABSI example, - how long they've had the central line, how long they've been in the hospital. Some specific impacts around that where a chronic condition that they've either already developed and you're trying to find out some sort of outcome afterwards. Yeah, you're going to need to look into kind of the history of other visits. Yeah, so I think that the use case is kind of following that spectrum of short term versus really long term.

Levi: Yeah, yeah.

Taylor: It's important to consider that.

Levi: Oh, totally.

And you see all the connection points between the machine-learning folks who are the analysts, because only the healthcare analyst can do this model creation themselves, and the connection with clinicians because I don't know everything

about CLABSI and COPD and CAUTI and all these different things. But when you're talking to a clinician or some sort of subject matter expert, they'll be able to say like, "Okay, well here are the things that might drive this prediction. And then with healthcare.ai, it boils it down and says, "Okay, from these 40 variables, here are the 20 that actually should be in the final model." So it's kind of a balance between clinical SMEE and algorithm.

Taylor: Yeah. And I'd say that something else that we're finding to be really important, as you consider some of these use cases at your health system or in your role or just things that you find interesting, is having some other sort of process or intervention around it, so using the using the predictions to drive some sort of improvement. So we're on a call the other day and someone kind of used the phrase "not doing machine learning just for the sake of machine learning". Well, there's definitely some interesting problems and that's kind of maybe where something like Kaggle comes in or you're learning it. That's really great. But then in actual practice, using machine learning to drive some sort of outcome or improve some sort of care is where you end up with the best adoption.

Levi: Yeah, exactly.

So go to Kaggle to practice. Don't necessarily practice your machine learning on your health system.

Taylor: Yup, you can definitely find something useful to do. There are so many awesome use cases out there. And a lot of people that are interested in having some additional support in their decision making so.

Levi: Exactly, exactly.

So we actually have a couple of comments that speak to some of those discussions.

Taylor: Oh, cool.

Levi: Yeah, so Dan Wellish, and this is a tiny bit of a non sequitur but it says, "Can you speak to your experiences with cleaning up data in terms of removing columns where imputing for nulls doesn't make sense?" That kind of goes back to the initial chat about, perhaps, grains. When you're rolling up to a particular grain, what if you don't have enough data in that column?

Taylor: Yeah. Yeah. I would think that you would want to profile the data pretty carefully up front. Yeah, if there are so many nulls that imputing is just not the right answer then, yeah, I think you'd want to drop that column.

And I think, in a lot of the clinical data, if maybe a certain field is pretty rare to be populated then maybe that's not a good one to consider. But you would also want to be careful not to toss out a field just because it looks like it has mostly nulls because that could be a really important event. And it is truly just as rare as you think it is or as it looks in the data. So you definitely would want to maybe consult someone else that has more experience with the process that's actually populating the data before you throw it out entirely, but it could be negatively impacting your model for sure.

Levi: Yeah, that's a great point. And healthcare.ai comes with tools that help you see, "Okay, what percentage of my columns or null? Do I need to actually fill in these null values with something that, you know, maybe the column mean or the column's most frequent value" - that sort of thing? So we've tried to kind of guide you along that process, what the tools are in healthcare.ai.

So a couple of other comments here coming through. So we had one other that asked us is John. So, actually a similar question. So, a lot of questions about data preparation asking about the granularity of data and dealing with missing data. So keep those coming in. That's super helpful.

And let's see here. Oh, okay. Thomas asked, "Do you know a good resource for a catalog of longitudinal data or patient-level data in healthcare?" So that's one of the hard parts of machine-learning in health care is that data availability is a little bit— well, it's difficult [inaudible 00:24:13] anything.

And with healthcare.ai, we actually put in a thousand-row longitudinal data set that is it's made up but it has pretty realistic values in it. And so, you can play with it, run the examples on it, and kind of get an idea as to, "Okay, well, if we have multiple rows per person, how do the algorithms interact with that?" So check out that built-in data set with healthcare.ai.

Taylor: Yeah. And that would probably be a good use, like the more longitudinal data [inaudible 00:24:41] have a good use for the linear mix model with—

Levi: Yeah, exactly.

Taylor: We've got some data for those patients where they've had multiple visits in your data set but it's the same patient. You might want to kind of aggregate those. But yeah, some of that really specific, real data is hard to come by though.

Levi: Yeah, maybe we should do an episode on that.

Taylor: Yeah.

Levi: It'd be kind of nice to browse around and see what's out there.

Taylor: Yup.

Levi: Yeah. It's definitely something where Kaggle can help a little bit. So check out Kaggle and see specifically their dataset page. But we'll be coming to you more with that in the future because that's definitely a hot topic.

Taylor: Yeah. That's a good thought, yup

Levi: Awesome.

So, Thomas here has a couple of other questions. And we appreciate y'all reaching out. Feel free to ask more. We're eager to interact here.

Okay. Thomas asked me, "I plan to add example datasets to test the functions in your R package?" And that's a great question. And so, yeah, we do have the one diabetes dataset. But, you know, we should put in a couple, I would say, because the small csv files don't take up much space. And maybe we could do something like a CLABSI dataset or a heart failure.

Taylor: Yup.

Levi: Hmm. Huh?

Taylor: Yeah, maybe try and work in some example datasets even towards some of the use cases that I mentioned earlier. But yeah, so obviously, it sounds like there's a little bit of a need for some longitudinal data out there. We could see if there's kind of a way to get at that a little bit to answer that question.

Levi: Yeah, yeah. That's an awesome point. So we'll follow up that and do maybe a whole episode on data out there that you can play with. Open data, we'll call it.

Taylor: Yup. And we're missing Mike today. I know that he has looked at some interesting data sources as well so we'll circle back with him.

Levi: Yeah. That's definitely a great point. We'll pass it on to Mike.

So, that's it. We went through the mailbag, talked about data and the news – Google bought Kaggle, and talked a little about models that you could build in your healthcare environment.

Just real quick, Taylor, what are some of the things that healthcare, you know, health systems are most concerned about? So if you think about, “Okay, well we can do models on all these different things. Where should the health system start? You know, maybe CMS? Just kind of pushing them in a certain direction?”

Taylor: Yup. I feel like re-admissions is definitely an area that health systems are focusing on. And it's also a tough problem. It's a tough problem to predict. There's a lot of variables that go into it that maybe are not available in just the EHR. Sometimes, the socio-economic data that's a little bit harder to get at is quite predictive. Anyways, I think that's a big area that we're seeing demand in that is good to focus on.

And then things like sepsis or some of the hospital-acquired infections. That's another one of those areas that impacts reimbursement and negatively impacts patient quality of life. And hospitals are starting to focus on all of the above now which is exciting for patients. But also, challenging for data scientists, data architects, data analysts. So there's a lot of interesting challenges out there.

Levi: Yeah. And there's a lot of things affecting reimbursements. And is this in an ACO sense you're saying or just hospitals in general?

Taylor: Ah, even in the hospital. Yeah, it can affect whether— you know, if there was something that you acquired while you're there that can affect an insurance or especially something like a Medicare reimbursement.

Levi: Interesting.

Taylor: And so, that's an area where if it impacts revenue and patient quality of life, the hospital system is going to start to look into it or should.

Levi: Yeah, that's a great place to start. That's awesome.

Now, thanks so much for joining us. Do we have anything else we need to address?

Taylor: No.

Levi: [inaudible 00:28:26] we kind of ran through our points there?

Taylor: Yeah, yeah. I think we kept things on track today. So, if there's any other things in the chat window, we could monitor that for a short bit. But otherwise, I think we got through most of the questions here.

Levi: Yeah. We'll take some for next week. And be sure to keep sending those in through email. We love to hear what you're up to.

Thanks for joining us, Taylor.

Taylor: I appreciate you including me today, so thanks for the opportunity.

Levi: Oh, for sure. Thank you.

Taylor: All right.

Levi: We'll see you all next week. Thanks, everyone.

Taylor: All right, thank you.