

**Narrator:** Welcome to the healthcare.ai weekly live broadcast with the Health Catalyst data science team where we discuss the latest in machine learning topics with hands-on examples. Here's your host, Levi Thatcher.

**Levi:** Hi, everybody. Thanks for joining us. I'm Levi Thatcher. This is the hands-on healthcare machine learning broadcast. I'm joined by Griffin Hoopes and Mike Mastanduno.

Mike, go ahead and introduce yourself. It's been a couple of weeks.

And then we'll have Griffin say a few words.

**Mike:** Yeah. So thanks, Levi. It's great to be on the show again.

My name is Mike Mastanduno. I'm a member of the Health Catalyst data science team. And, I guess, my main job has been building healthcare.ai and helping people use it so it's been a lot of fun.

Griffin.

**Levi:** Special guest star.

**Griffin:** Yeah. I'm Griffin Hoopes. I've been a catalyst for over a year. I've worked in healthcare for over 10 years now and that sounds really weird.

**Levi:** That's impressive. That's longer than me.

**Griffin:** Yeah, I've been exploring machine learning for the past few years and have been really interested in it. And with healthcare.ai it's made it easier for me. Yeah, it's exciting.

**Levi:** Glad to hear. We didn't pay him to say that. No. We didn't [inaudible 00:01:32] him even.

So to start off, just some logistics. We're super excited to have you on, Griffin.

So be sure to log in to your YouTube account if you want to chat. And we encourage you to chat and implore you to chat because that's what drives the discussion. And we want to hear what your problems in machine learning and healthcare. And in case you want high resolution you can play with the little gear there in the bottom right corner of your YouTube window and increase the resolution.

Don't go too close. If you heard about these high-resolution movies and what not, sometimes it's a little too close. So if you get too much resolution, you may want to back down. But you can toggle that up and down if you want.

And then subscribe. We'd love to have you with us each week, you know, share, like, whatever.

But we're going to start with the mailbag today. And we've got a couple of different questions coming in this week. So this is a broad one. This is a goodie. "What is the difference between machine learning and artificial intelligence or AI?" AI has been in movies. You've heard about since the 50s but—

Mike, any thoughts?

**Mike:** Sure. Yeah. I mean, they're both kind of hand-wavy terms. I think that can mean a whole bunch of different things. I think, in general, kind of machine learning would capture building models that learn from data and using them to predict things. So healthcare.ai - that's machine learning. That's what we do.

And then using past data to predict future length of stay. That would be an example of machine learning. Whereas, artificial intelligence might be something like **how** from that one movie or—

**Griffin:** 2001?

**Mike:** Yeah, that's the one.

**Levi:** Yeah. Exactly.

**Mike:** Thanks, Griffin. You see, we knew we brought him along for a reason.

And getting into things like maybe you could consider a self-driving car to be artificial intelligence because presumably it's going to get better at driving itself over time—

**Levi:** Yeah, yeah.

**Mike:** And it's going to be learning from experience, whereas standard machine learning might not do that.

**Levi:** Yeah, that's a great point. There's this podcast called Talking Machines. And the host, Ryan Adams, is this professor. And he put it this way. He said that, "Whenever we look forward and anticipate technology like self-driving cars, we

always call AI. And then when we get there we call it machine learning because basically it's just Math." So it's kind of a fuzzy definition but that that's the one way to think about it.

**Mike:** I mean you can get it even more complicated with deep learning.

**Levi:** Yeah, deep— well, that's—

**Mike:** Which is kind of a hybrid of the two.

**Levi:** Great segue into our second question, "What's the difference between regular machine and deep learning?"

Griffin, have you gotten into deep learning at all?

**Griffin:** I haven't. It's actually pretty different than just your usual machine learning. Neural networks - they can solve a lot more complex problems than your typical machine learning. They can do you know, image recognition, natural language processing.

**Levi:** That's a good point.

**Griffin:** You know, they even use deep learning to learn how to play video games and beat them, so it's a newer thing and it's more complex.

**Mike:** Yeah.

**Griffin:** It's definitely-- it has a lot of opportunities.

**Mike:** It's not quite as formulaic as like a tabular data set. You know, random forest is like you follow the same procedure for every model you build. But deep learning is like, "Well, if you're trying to translate language the way you built that model, it's much different than if you're doing something like driving a car.

**Griffin:** Right.

**Levi:** Yeah, yeah. On the sexiness meter, the deeper the learning, the sexier the learning.

**Mike:** Definitely.

**Levi:** It's kind of how it plays in the industry these days. So it's sort of deep on anything, you know it's marketing genius. Mike's getting into deep learning. Or, you know, neural networks is kind of a more familiar term.

**Mike:** Again, that's the—

**Levi:** A formal term, I guess, you could say?

**Mike:** That's the going on at healthcare.ai right now, is trying to get neural networks to work on our tabular data, just hoping that as these datasets get really, really big maybe we'll be able to get a little more information out of them using a neural network.

**Griffin:** Yeah, it's something that I've thought about with a neural network and healthcare is I had an eye test done last month. And they look at an image of your retina and they can tell if there's bleeding or any kind of problems. And I thought, "Well, that would be cool if you could take a bunch of retinal scans and you could teach your machine to show, "Hey, there's a problem here in your eyeball." And, you know—

**Levi:** That's an amazing use for this.

**Griffin:** Yeah, there's a bunch of things like that where eventually, I think, it will probably get to the point where computers are looking at those images and saying, "You need to follow up with your doctor. Oh, he's okay."

**Levi:** Yeah, yeah, yeah. That's an awesome use case. So there's something called diabetic retinopathy, I believe, and Google DeepMind has actually been looking into that. Apparently, diabetics have certain almost calcifications or issues with their eyes. And this model was able to detect a lot sooner than perhaps a clinician would. But obviously, it makes a lot easier on the clinician to have this alerting system set up. Thanks for the help with that. That was awesome.

And that's a good amount of questions. Keep them coming in the chat window so we can respond in real time as the broadcast goes on.

Should we dive in into this topic?

**Mike:** Yeah, that sounds great.

So, today's topic, we've advertised that we're going to do more of an in-depth case study of how someone has used healthcare.ai. And so we're going to do just that to talk about a model that Griffin built.

Griffin is the-- you know, we envy him a little bit because he actually gets to play with the clients' data. And we're kind of one level removed, so we have to work

through Griffin. Using healthcare.ai, we develop the tools and then we help him use the tools to actually impact patient care

So, Griffin, I think we could start with just, you know, what's the model that you built?

**Griffin:** So to give a little bit of background on the model that I built, I'm working on an application that we have here at Catalyst called Patient Flow Explorer. And what it does is it analyzes a patient's journey through the hospital. And something that is tied very closely to that is their length of stay. It's a big issue at hospitals. Patients taking up bed space and you have to get people out as you're getting people in. So it's a big concern for hospitals.

And so, what I ended up building was a classifier that could predict if a patient was going to spend more than five days in the hospital. And you can pick and choose and change that number how you want. But how I came to that five days numbers is that I looked at our clients' data and I saw that the average length of stay was around five days and I thought, "Okay, well, anything over that, I want to flag and alert the client so.

**Levi:** That's fantastic.

**Griffin:** Yeah.

**Levi:** So number one business problem, right? Would you say?

**Griffin:** Mm-hmm. Yeah.

**Levi:** What do you need to solve?

**Griffin:** So what you need what you need to solve this problem. If you think about it, you might not know a whole lot about the patient when they come to the hospital. You know how they arrived. You know their age. You know their gender. You might know a little bit about their history.

You might have an admitting diagnosis if a doctor makes the decision to admit them and says, "You know, you've got chest pains. We're going to diagnose you with heart failure and admit you into the hospital." So you only know a certain amount of things about the patient before you know really how long they're going to stay in the hospital.

**Levi:** Okay.

**Griffin:** So yeah. It's very basic information but on our clients' data where we have years and years of hospital visits, we can learn. We can point the computer at it and we can say, "Hey, go and learn these patterns. Are older patients - do they end up staying there longer? Do females versus males, do they stay there longer? Does one type of diagnosis versus another stay there longer? If my patient arrives in the hospital on a helicopter versus—

**Levi:** That might—

**Griffin:** --a taxi.

**Levi:** --affect the outcome.

**Griffin:** Yeah. Absolutely

**Mike:** That's [inaudible 00:09:54] I guess.

**Griffin:** It's like, yeah.

**Levi:** Social determinants of health—

**Griffin:** Yeah.

**Levi:** Billionaires, you know, for hospital stays.

**Griffin:** Well, yes, there's that but— yeah. I'm thinking more about life flights. You know, if you're coming in with a gunshot wound or something.

**Levi:** That's a— yeah.

**Griffin:** But yeah, I mean, you could be a billionaire who shows up in your helicopter and just wants to stay a week in the hospital and—

**Levi:** Yeah and think about—

**Griffin:** --eat the cafeteria food.

**Levi:** --sort of, medivac entrances. They're like common. That's a really interesting data point. So, basically, you're gathering and seeing what features could help predict something?

**Griffin:** Mm-hmm.

**Mike:** So just before we dive too far into the model, I just want to make sure that we're framing our conversation around the standard machine learning pipeline we've

been talking about on the show. And so, that pipeline is something like identify a business question and you've done that, you want to measure length of stay to help your client be able to forecast who's going to need a bed for help longer than the average.

And then you're going to have to get a data set. You're going to build a model. You're going to do some feature engineering and then evaluate that model.

And finally, you're going to move it into production where it's making predictions every day. So we'll revisit that timeline just for our own grounding.

But before we get into the data you chose, you used healthcare.ai to build this model, right?

**Griffin:** Yeah.

**Mike:** That's great. What was the installation process like?

**Griffin:** Well, before I talk about the healthcare.ai installation process, I actually used my own model. I didn't use healthcare.ai. I got data out of the database. I put it in a CSV. I transformed the CSV. I started experimenting with the different classifiers and just seeing what would work. And I was thinking—

**Mike:** And you used Python, right?

**Griffin:** Yeah, I used Python for it. And healthcare.ai has a Python package so I was thinking, “Wow, I've put in a lot of work on this and this isn't going to be very scalable. It's like, how would I how would I go through this whole process and repeat this every client that I go to? This isn't going to work.” And so I was seeing the messaging from our data science team and saw that healthcare.ai, you know—

**Levi:** We've sent out a lot of spam.

**Griffin:** There is. And it comes to in your email and I said, “Hey, you know, this looks great. They've got a Python package here. They also have an R package.” I have a background in Python development. And so, I downloaded the Python package.

We actually have really great documentation for healthcare.ai. I watched on my own. And yeah, it's a credit to you guys. If the documentation isn't good, you're just going to be stuck there scratching your head and figuring it out. So it really walks you through how to connect to a database.

And it makes a lot easier. So I didn't have to take the data and put it in CSV. And then transform that. It says, "You can just query your database on your own." So yeah, actually, working with healthcare.ai model, it simplifies it. It allows a user to—

I mean, right now, we can pick between, do you want to do a regressive model which an example of regression would be-- and this, to me, would be the Holy Grail to say, "When Levi checks into a hospital, I would love to be able to produce a number that says, he's going to be here for 40 hours." And I would love to actually put a real number on each patient that comes in. And I think--

**Mike:** Definitely something to work towards.

**Griffin:** Yeah, I think that we can get there.

**Levi:** Probably amazing. Depressing to me but hopefully the clinician--

**Griffin:** Hopefully, you don't have to spend 40 hours in the hospital but—

So, I had to choose between a regression model and a classification model. And so, the classification model is a little more simple, for me, where I could say, "Is Levi going to stay in the hospital for more than five days, yes or no?" So to me, that that helped kind of simplify the problem so that I can approach it, you know?

**Mike:** Yeah.

**Griffin:** I wasn't going for the "Levi is staying for 40 hours" because that's starting at the hardest point.

**Mike:** Yeah and that's a great point. Like, a lot of times it's about framing your question in a way that's easy to answer - or easier to answer so that you can work towards a more difficult solution.

**Griffin:** And knowing that I have that data available, I think I have the data where I can classify Levi into a bucket - yes or no.

**Mike:** Usually, we use Joe around here but—

**Griffin:** Joe.

**Mike:** But, I mean.

**Levi:** Yeah, well—

**Mike** Levi works too, I guess.

**Levi:** Whatever works. So did that come across from like a clinician's prospective? I wonder if five days is like a critical point? Like, maybe we need to flush out from their side?

**Griffin:** Well, as I said earlier, the five days number came from just doing a quick analysis on the client's data to say, "The average length of stay per patient was at five days." And so, I figured, anything over that is abnormal and we'd want to flag that.

**Levi:** Yeah.

**Griffin:** But yeah.

**Levi:** It's a good start.

**Griffin:** It's a totally configurable number. When you work at your hospitals, they might say, "No, we want to know - after 24 hours, who's staying on there?"

**Levi:** Yeah, exactly, just the business case kind of dictates.

We did have a question coming through here. So somebody asked about, "Did you see any correlation with discharge disposition?" Do you know that field in the EMR, discharge disposition?

**Griffin:** Okay. So, discharge disposition, we don't know what the discharge disposition is going to be when somebody checks into a hospital. That's—

**Levi:** That's hard.

**Griffin:** Yes, so—

Well, I mean, we have to predict the future for that. I mean, we can look at the past.

**Mike:** Right, that's what we do though. We predict the future.

**Levi:** So, how did you-- did you say, after 24 hours, that's when we're going to make the prediction and whatever variables are available then we'll use or—

**Griffin:** Yeah.

**Levi:** How do you make that decision?

**Griffin:** So based on how our data comes in, there's a 24-hour refresh. And so, most of the time, you're going to get a window of the patient will have already been admitted for 12 to 24 hours before I'm actually getting the data in. Eventually, it would be really cool to do it in real time - we're very close to real time, where I say, "Right when you check in, we can make a prediction about how long you're going to stay here" but just the nature of the data that I was working with. I am—

**Mike:** That's a hard technical.

**Griffin:** --working in a development environment, on our client's side, so there's a bit of a lag so.

**Mike:** We barely got off paper so going to real-time data logging is—

**Griffin:** Yeah.

**Mike:** --that's quite a jump.

So maybe you could talk a little bit about the data set that you had. You know, you grab data from the electronic health record system and about the patient demographics, the social info. Like, how much of it? What kind of sources did you pull from?

**Griffin:** Yeah, so I have years of data. I didn't want to go back too far. I wanted the data to feel more recent because the client that I'm working with actually has a newer facility that they've opened.

**Mike:** Oh, okay.

**Griffin:** So I wanted to, I guess, skew that date more heavily towards how that campus is performing. So I ended up having about around 50,000 encounters that I analyzed to determine the patient's length of stay. And like I said, I looked at things like the acuity level of the patient – meaning, how many resources are assigned to them when they come into the hospital. If you have a higher acuity you've got, potentially, five nurses working on the gunshot wound versus maybe an acuity level of one where maybe you've broken your wrist or something. And so—

**Mike:** Yeah.

**Griffin:** So I looked at acuity. I've looked at the admission source. Like, where—

**Mike:** Like, helicopter versus—

**Levi:** Sorry, how often is helicopter? Like it's, you know, percentage of all—

**Griffin:** It's—

**Levi:** [inaudible 00:18:05] pretty rare?

**Griffin:** It doesn't show up a lot.

**Levi:** Okay.

**Griffin:** But it's enough in there that it makes a difference.

**Levi:** That interesting to hear. That's fascinating

**Griffin:** Yeah. It's important for a feature. Of course, I looked at patient's age, their gender, their admit diagnosis.

**Levi:** Oh, yeah.

**Griffin:** So yeah, a lot of— yeah, it's like when you're admitted – yeah, that can help. It's like it can definitely help.

**Mike:** Yeah, so kind of after you've framed your business question and gone kind of looking for the data like someone in your position is really in a great spot to build the right data set because you have so much familiarity with the data and that's the real creative part of machine learning. And, you know, the algorithms are available to anybody. But to build a really good model, you need good data so.

**Griffin:** And, for me, it's also a great opportunity to collaborate with people. I collaborated with really smart people here at Catalyst and I said, "You have experience in healthcare, what would you use to determine how long a patient would stay?" And that's where I got some of those ideas.

Of course, I had my own ideas, like "Maybe, I can check how their how their blood pressure measured when they came in? What was their weight? What's their employment status? Maybe if they're insured versus not insured, do they stay longer?" So I had a lot of guesses myself and I tested all of those theories and measured the importance of those features. And I also had to weigh that against our expert advice.

And I read a couple of papers based on other people's studies and government studies about the contributing factors to length of stay. So yeah, it's kind of

important to also note that it might not necessarily translate to every client or hospital that you have.

**Mike:** Yeah.

**Griffin:** One factor that might be really important for one client might be meaningless at another so.

**Mike:** That's kind of the beauty of healthcare.ai.

**Levi:** Exactly.

**Mike:** It's like you can customize it to the specific department, the specific hospital, geographic region or whatever you want.

**Griffin:** Mm-hmm.

**Levi:** Yeah, that's beautiful. And then the person most familiar with the data is actually the one creating these features and doing the modeling and machine learning.

We did have another question come in. So somebody asked about using this for more kind of a top-level view. So you kind of mentioned in the sense of, "Okay, maybe clinicians would want to know who's going to stay the longest in my department". But maybe the hospital itself, in terms of a census count, how many people going to be discharged over the next 24 hours or the next three days? I guess, you can kind of like summarize it to have a whole different perspective?

**Griffin:** Yeah, you could. You could.

I think, maybe a more interesting question to ask around census would be actually predicting the census count every night and say--

**Levi:** There you go and we're forecasting maybe?

**Griffin:** Yeah.

And I think that's a different problem solve. But, I mean, you could take the answers from this length of stay and kind of have that top-down look to say, "Okay, I'm expecting this percentage of patients to be discharged. I'm expecting this percent to be taking up bed space." But yeah, when it comes to census, I would think more about actually predicting the census—

**Levi:** Just the count of people in each department over a certain number of time or a month.

**Griffin:** And it's a big issue at hospitals.

**Levi:** Because they're full up. Like, people don't want to stay there for very long. And so, [inaudible 00:21:49] get them out quickly. Interesting.

So yeah, we're kind of running out of time here. Any final wrap up points, maybe?

**Mike:** Well, let's see. Can we give it like a real quick what was the model you've ended up using and how did it do?

**Griffin:** So I ended up choosing the linear model. I had to choose between that and a random forest. And the linear model was performing a higher AUC score. So that's the one I've gone with. I'm testing it in the wild now. And the client's production environment is very similar to their development environment so I'm expecting good results when we deploy it to production.

**Mike:** Cool. And so, where's the model now? It's in the wild testing, right?

**Griffin:** It's in the wild testing. It's been out there for a few weeks and it's producing a pretty solid score above 0.8 AUC right now.

**Mike:** AUC.

Okay. That's awesome. So, I guess, after it comes back, we're going to hopefully see that high performance continue. And then how's that information going to end up getting served up to the to the clinicians who might use it?

**Griffin:** That's a good question because we want to visualize that data when we're done with it. So, as I started out saying I was working on an application called Patient Flow Explorer, I think that I can end up putting a visualization in that application to say—

**Mike:** Oh, okay.

**Griffin:** -- "Here's a look at a list of patients and their probability that they're going to be in the hospital for more than five days." Possibly have visualizations of what the most important features are—

**Mike:** Okay.

**Griffin:** --to make those predictions. Yeah, I think that there's opportunities to retrofit and embed the machine-learning classifications into our applications.

**Mike:** Well, that's great. That's so exciting that you were able to do that work and kind of get it into production. I just have two more questions for you.

**Griffin:** All right.

**Mike:** And these are the big money questions. So would you do it all again? Would you build another model?

**Griffin:** Oh yeah, absolutely.

**Mike:** Awesome. That's so great to hear.

**Griffin:** Yeah. I'm actually working—

**Mike:** I swear, we didn't pay him.

**Griffin:** I'm actually working on a couple more that I'm testing in parallel.

**Mike:** That's great to hear. And how did you find the healthcare.ai experience, you know, from your position to deploying the models?

**Griffin:** It simplifies it. It makes machine learning portable and scalable, I think. It's really easy. It's kind of-- you just plug it in and you connect your data to it, and it kind of does all the work for you.

**Mike:** Man.

**Griffin:** It is a lot easier than doing it by myself, like how I started.

**Mike:** I couldn't have gotten a better response if I'd asked him again.

**Levi:** No, but he came to us. So we didn't even pay him to be out here.

Awesome work, Griffin. That's really cool work.

**Griffin:** Thanks.

**Levi:** And we're excited to see, you know, where do we go from here? Maybe we'll be able to follow up in a few weeks to see what's happened since?

**Mike:** Yeah, that would be cool.

Yeah. So, I guess, we're out of time for today but next week we're going to be-- Levi's out, so I'll be running the show, and we'll be talking about tips and tricks for high performance models in the wild.

**Levi:** Oh, I'm going to miss that one.

**Mike:** Which is definitely a sticky topic, you know?

**Levi:** That's right.

**Mike:** It looks like I got stuck with that one a bit.

But please, again, subscribe to our channel. Join the community of healthcare.ai users. We'd love to have you come back. Give us more questions. The questions have been really great this week. And we hope to address them again at the beginning of the show.

So, do you guys have anything else to add?

**Levi:** No, thanks guys.

**Griffin:** Thanks, everyone.

**Mike:** Thanks so much everybody.

**Levi:** See you next time.

**Mike:** Thanks, Griffin.

**Griffin:** Yup.