



chr

SDS PODCAST EPISODE 755: BREWING BEER WITH A.I., WITH BEAU WARREN

Show Notes: <http://www.superdatascience.com/755>



- Jon Krohn: 00:00:00 This is episode number 755, with Beau Warren, president of Species X Brewing. Today's episode is brought to you by CloudWolf, the Cloud Skills platform.
- 00:00:13 Welcome to the Super Data Science Podcast, the most listened to podcast in the data science industry. Each week, we bring you inspiring people and ideas to help you build a successful career in data science. I'm your host, Jon Krohn. Thanks for joining me today. And now, let's make the complex simple. Welcome back to the Super Data Science Podcast. Today, we've got a one-of-a-kind episode for you in which we hear about how one of our listeners reached out to me to get my help with creating an innovative beer suggested by AI. That listener, Beau, is the head of brewing at Species X in Columbus, Ohio, where he's using machine learning to brew groundbreaking beer. He has a decade of previous experience working in brewing roles. He is a Certified Level 2 Cicerone, meaning Certified Beer Expert. He studied microbiology at the Siebel Institute's Montreal campus and he has a certificate in data analytics from Ohio State University. He was also a star college football player at Virginia Tech. He was invited to play football professionally with the NFL's St. Louis Rams, but opted to focus on beer and AI instead.
- 00:01:29 Today's episode will be of interest to anyone who loves beer or dreams of applying AI to their area of domain expertise, regardless of whether you're already an avid data scientist or an aspiring one. In this episode, Beau details how beer is brewed, how Beau moved from being a listener to this podcast to a collaborator with me on his AI beer project, how he wrangled years and years of brewing data to create a super-valuable proprietary dataset for training machine learning models, how he leveraged his self-taught ML skills to generate beer recipes humans wouldn't have conjured up themselves. He talks about the specific Python libraries he used to train his ML models



and select the beer recipe that's most likely to be a big hit with beer drinkers. And, we reveal the name and the style of our first AI beer, as well as when and where you can try it yourself. All right, you ready for this delicious episode? Let's go.

00:02:28 Beau, welcome to the Super Data Science Podcast. It's so cool to have you here. For people who are watching the YouTube version, they'll see me for the first time ever on the podcast, I'm having a beer on air. There's good reason to be doing that, because you're a brewer. Tell the audience how we met, Beau. How did this all come about?

Beau Warren: 00:02:55 I was walking my dog, I remember that specifically. I'm trying to remember, I think it was in 2022? Maybe it was 2023. I already had this idea to really utilize AI and data analytics to create beer. I already started on the project. I was listening to episode, I forget the episode. I know you'll be able to help me out with this one.

Jon Krohn: 00:03:28 It was the episode about beer, so it was the episode with Zack Weinberg on using ChatGPT for-

Beau Warren: 00:03:35 Correct.

Jon Krohn: 00:03:36 It was the layman's guide. Then, the context here is it was definitely, it was early 2023. That episode came out on January 20th, 2023, it's episode number 646. In that episode, my friend, Zack Weinberg, he's one of my oldest and dearest friends, him and I had been talking. His mind was so blown about ChatGPT, which at that time was brand new. He was using it for tons of purposes within his brewery to make it easier to create copy, to create marketing materials, to answer questions in an automated way that people might have, particularly on his website. Yeah. He came on the show to do this episode, it was a Friday episode, on ways that non-

programmers, laypeople I guess, could be taking advantage of AI tools like ChatGPT to automate things. Yeah, I guess that's the episode you're referring to, that was about a year ago, episode 646.

Beau Warren:	00:04:42	Yeah. I was listening to it and I was like, "I'm literally doing that." Because I know he mentioned, "Someone should get-"
Jon Krohn:	00:04:52	Right.
Beau Warren:	00:04:53	"All these feature columns together."
Jon Krohn:	00:04:54	That's right.
Beau Warren:	00:04:54	You know what I mean?
Jon Krohn:	00:04:55	Yeah, that's right.
Beau Warren:	00:04:57	I'm literally doing that right now.
Jon Krohn:	00:05:01	Yeah. I think his idea in the episode, if I remember correctly, he said something like ... Because, at that time, ChatGPT 3.5 was trained on everything on the internet, there's tons of brewers that just publish their recipes. He was like, "You could be using the knowledge embedded in this large language model, in order to get recommendations on how to brew a recipe." I don't think it was as specific as what we're going to get to later in this episode.
Beau Warren:	00:05:27	Oh, okay.
Jon Krohn:	00:05:27	With feature columns.
Beau Warren:	00:05:28	Gotcha.

- Jon Krohn: 00:05:30 But, definitely he talked about the ideas of ways you could be using LLMs to generate beer recipe suggestions. If I'm remembering correctly, yeah.
- Beau Warren: 00:05:39 No, I think you're right. I was doing similar, I guess. Not literally that. I was like, "Man, I got to reach out to Jon." I remember I hit you up on Twitter, crossing my fingers. You hit me up and you're like, "Dude, let's chat." I was like, "Let's." Yeah, it just went from there. One thing led to another, and all of a sudden, we're doing a collaboration beer together. I'll tell you what, I'm really excited how it's come out.
- Jon Krohn: 00:06:22 Yeah. We're going to do the big reveal later. To give an overview, we started working on this collaboration. Or, we started talking about potentially collaborating a year ago. That picked up in the fall, the Northern Hemisphere fall of 2023, as you started to get a new brewery together called Species X. I don't know if you want to tell us a little bit about that, because that's also ... People, even our audio-only listeners, will be able to hear the rich echo of this massive space that you're brewing in. People watching the YouTube version can see this epic background behind you.
- Beau Warren: 00:07:02 Yeah.
- Jon Krohn: 00:07:03 Do you want to tell us about Species X Brewing?
- Beau Warren: 00:07:06 Yeah. Species X started out in, I think it was 2021, at least the idea came around. The whole idea was ... This was before the pandemic started, I had this idea. I listen to a lot of scientific podcasts, so one was about microbiology. They were talking about finding the next pathogen. How do we prepare for the next pathogen?
- Jon Krohn: 00:07:37 That was before COVID?



- Beau Warren: 00:07:38 That was before COVID. They called it Pathogen X. It was like, "I would love to find the next great organism or species in beer." That's how I came up with Species X. Now granted, AI's a little tongue-in-cheek. People say that could become the next species, so I'm kind of taking that, not super seriously, but as part of the brewery. Then, the other part of the brewery is all novel biological organism-focused. So genetically modified yeasts, force mutated yeasts, wild captured yeast, spore germinated yeast. Just all things biological and novel in nature, at least to humans. We have those two separate realms. We call the Silicon Species is the artificial intelligence, ML models that I've created and coded. Then, the other is the carbon species, so that's going to be our GMO yeasts, novel biological organisms.
- Jon Krohn: 00:09:03 Gotcha, yeah.
- Beau Warren: 00:09:04 Yeah.
- Jon Krohn: 00:09:04 Species X, I've even seen that on your website, when people go to buy beers, which will be easier and easier. Right now, it sounds like there's only a couple of beers available at the Species X website, and those are maybe at the time of recording, only available for shipping in Ohio.
- Beau Warren: 00:09:25 We do not ... Those four beers online are our collaboration beers. They're only available at Aslin.
- Jon Krohn: 00:09:35 Oh. Where you used to work. I see, got it.
- Beau Warren: 00:09:38 Yeah. We will have four flagships. Four flagships will be available through distribution in Ohio. We have an alternating proprietorship brewery in Virginia that will be brewing that for distribution. Then, we'll be doing parlor batches here, so little three-and-a-half barrel batches of beer, which equate to seven full kegs of beer.



- Jon Krohn: 00:10:08 All right. What you're saying is that, at the time of recording, there aren't real Species X beers brewed at Species X that people can yet buy online, but it's going to be happening more and more as the brewery opens more and more. It's something that's just happening. You've been kicking off the Species X Brewery. The main thesis that you have, as president, and I guess head brewer of Species X?
- Beau Warren: 00:10:34 Yeah, head brewer. I'm trying to think of all the hats I'm going to be wearing in here. I'm the only production person, president. I'm taking care of costs, managing all of our employees. Looking at costs, and P&Ls, that sort of thing.
- Jon Krohn: 00:10:59 Right.
- Beau Warren: 00:11:00 Really exciting stuff.
- Jon Krohn: 00:11:02 Yeah. For the thesis there, as president, and manager, and P&L head, and production person, all these different hats that you're wearing at Species X, there's two main theses. One of them is that novel biological organisms, like GMO yeast, can produce interesting beers in a way that no existing yeast could or no other brewery could. You call those your carbon species?
- Beau Warren: 00:11:32 Yeah, correct. Yeah.
- Jon Krohn: 00:11:33 Then, you also have your silicon species, which is mostly what we're going to focus on in this episode, or eventually we'll get to focusing on quite a bit.
- Beau Warren: 00:11:40 Correct, yeah.
- Jon Krohn: 00:11:43 You've written a ton of code. What's the word? I've seen the code. I haven't worked on it myself, I haven't written any code, but I've reviewed some of your code and had



some ideas of ways that you could make changes, potentially to data that you're including or ways that you could be processing or modeling the data. We're going to talk about all of that silicon stuff later in the episode. We're going to talk about how you have been using machine learning in those silicon species to create what could be some of the greatest beers ever. And, I'm going to get to find out, on air, how our first beer is coming along. We're going to get to that later in the episode.

00:12:31 But first, let's talk about beer in general. You are a certified, literally certified, beer expert. You're something called a cicerone. So people who go to fancy restaurants, they may be aware of the job title sommelier, which is somebody who is a specialist in wine tasting. The equivalent in the beer world is a cicerone. I've only read about this in The Economist. I don't know a huge amount about cicerones. But I know that you are a Level 2 Cicerone, so maybe you can tell us a bit about that. Another big credential, other than your many years, you have about a decade of experience working in brewing, but you also studied microbiology at the Siebel Institute in Montreal. Tell us a bit about how somebody becomes as expert at beer as you, and what these kinds of certifications, like the cicerone mean, and why it would be useful to study microbiology as a brewer.

Beau Warren: 00:13:29 Yeah. Obviously, I'm obsessed with the art of brewing. In addition to the art of brewing, I'm absolutely obsessed with science. This was back before I stepped into full-time production mode, I worked at sales for several companies. One of these companies allowed us to study to be a cicerone through the cicerone program. We had a pretty intense two weeks, studying that. Level one is Certified Beer Server. You can actually take that online fairly quickly. It's still very hard. You definitely need to study for it. Level two is exponentially harder than that. I

believe, from what I remember, it's about a five to six-hour test. There's a tasting.

Jon Krohn:	00:14:27	A five to six-hour test. Do you get a bathroom break?
Beau Warren:	00:14:31	I don't know if I'm right about that. I'm trying to remember, because it's so long ago.
Jon Krohn:	00:14:36	You were drinking at the time.
Beau Warren:	00:14:41	There was a sensory portion, which didn't take that much time.
Jon Krohn:	00:14:44	Oh, really?
Beau Warren:	00:14:44	Yeah, a sensory portion.
Jon Krohn:	00:14:46	So level two's obviously in-person.
Beau Warren:	00:14:47	Yeah.
Jon Krohn:	00:14:48	That was implied by you mentioning the level one could be done online. Then yeah, if there's a sensory portion, then that must be done in-person.
Beau Warren:	00:14:55	Yeah.
Jon Krohn:	00:14:56	That's cool.
Beau Warren:	00:14:58	Then, there's a three-hour written portion. Maybe it's more like four to five hours.
Jon Krohn:	00:15:06	It doesn't matter, we're getting the rough idea.
Beau Warren:	00:15:08	Yeah, anyways. That was really hard. I found that there was some odd flavors that I just did not have the right threshold for it. I had to take the test twice because of the sensory portion.



- Jon Krohn: 00:15:23 Oh, really?
- Beau Warren: 00:15:23 Yeah.
- Jon Krohn: 00:15:23 You aced the written part, but the sensory part was hard.
- 00:15:27 Data science and machine learning jobs increasingly demand cloud skills. With over 30% of job postings, listing cloud skills as a requirement today, and that percentage set to continue growing. Thankfully, Kirill and Hadelin, who have taught machine learning to millions of students, have now launched CloudWolf to efficiently provide you with the essential cloud computing skills. With CloudWolf, commit just 30 minutes a day for 30 days, and you can obtain your official AWS certification badge. Secure your career's future, join now at CloudWolf.com/SDS for a whopping 30% membership discount. Again, that's CloudWolf.com/SDS to start your cloud journey today.
- Beau Warren: 00:16:09 Yeah, correct.
- Jon Krohn: 00:16:10 That's amazing because obviously, assuming that you did this certification relatively recently, you already would have had a decade of experience professionally tasting beers. That really says something about this tasting.
- Beau Warren: 00:16:22 No. I was in the industry for probably three years by then. Like I said, this was pretty far back ago. Some people just don't have a really high threshold for things. There was one or two things that I just had struggled picking up. Anyways, I got that, which I was really excited about. Beyond Level 2, there's also Advanced Cicerone, and then there's Master Cicerone. Advanced is also exponentially harder than Level 2.
- Jon Krohn: 00:16:59 Oh.

Show Notes: <http://www.superdatascience.com/755>



- Beau Warren: 00:17:02 Then, after Advanced Cicerone, that's exponentially harder than that. I have nothing but the most respect for not only the people taking Level 1 and 2, but seeing how I failed the first time on getting Cicerone Level 2, how hard and dedicated you have to be to achieve Advanced Cicerone.
- Jon Krohn: 00:17:30 Yeah.
- Beau Warren: 00:17:30 And to get the Master Cicerone, which I think there's only a handful of people in the world that have Master Cicerone.
- Jon Krohn: 00:17:37 That must have been what I was reading about in The Economist. It's dozens of people, worldwide, that have it.
- Beau Warren: 00:17:42 Yeah.
- Jon Krohn: 00:17:44 Yeah, that's wild. There must obviously be some kind of ... Maybe not obviously. There might be, I suspect, a strong biological component to being able to do this. My sister, for example, she can just taste way more stuff than me. I've never been a great taster. We actually looked into this, growing up. There's a density of taste receptors on your tongue and people have different densities. You can have these little, they're the size of a Cheerio, a flat piece of paper that looks like a little Cheerio. You can put that on your tongue, and you can count how many taste receptors are in that little Cheerio-looking circle thing.
- Beau Warren: 00:18:26 Really?
- Jon Krohn: 00:18:27 Yeah. That gives you a sense of how many taste receptors you have on your tongue. Some people have way more than others.
- Beau Warren: 00:18:34 That's fascinating. I did not know about that. Start handing those out.



- Jon Krohn: 00:18:40 Yeah, you could. You could. You could look that up and definitely, it's something you can buy. It could be interesting, actually, if people are coming by and doing a tasting at the brewery. You could have these there. It just takes a minute to count, and then it gives you a sense. There's super tasters, which my sister is because she has a crazy number of taste receptors. I'm, I can't remember what the word is, for a sub-taster. I'm like deficient.
- Beau Warren: 00:19:09 It's funny, my wife is the same way.
- Jon Krohn: 00:19:13 She's a super taster?
- Beau Warren: 00:19:14 Yeah, super taster. She's really good.
- Jon Krohn: 00:19:16 You'll have to do a test to see whether it's experience or biology in her case.
- Beau Warren: 00:19:20 Yeah.
- Jon Krohn: 00:19:23 Cool. Yeah, that's the cicerone. Then yeah, in addition, you studied microbiology. We're going to get into how beers are made in a moment. But, I guess people might not appreciate the extent to which brewing beer involves working with living biological organisms. Microbiology training can also be a huge asset. Do you know many other brewers that actually formally studied microbiology like you did?
- Beau Warren: 00:19:50 I'm trying to think, here. Obviously, the bigger brewers around are going to send off the right people to go get trained. Obviously, all lab techs, you definitely want some sort of lab training, which is prominently why Aslin sent me for two weeks in Montreal to go study beer microbiology.
- Jon Krohn: 00:20:15 Right.

- Beau Warren: 00:20:16 I was the head of the lab, so I did all the QAQC, microbiology checks, all that stuff.
- Jon Krohn: 00:20:23 Yeah. I'll explain a few terms here. Aslin, that you've referred to a couple times now in the episode, that's a brewery in Alexandria, Virginia, that you worked at for five years. When you say QAQC, this might actually be obvious to some software people out there because QA is also software engineering specialization, it's the quality assurance. The QAQC is Quality Assurance and Quality Control supervisor you were. Yeah, I see. As part of that role, you're spending a lot of time ... Are you literally working with test tubes and that kind of thing? Yeah?
- Beau Warren: 00:21:00 Yeah. I was making plates. We had a nice set of selection for our plates.
- Jon Krohn: 00:21:11 Plates for growing a bacterial culture kind of thing?
- Beau Warren: 00:21:13 Exactly, yeah. What I love to do is ... My favorite plate was WLN because it would change colors depending on the yeast strain, the species of yeast, the bacteria, all that stuff. It made identifying colonies a lot easier. That was my go-to general media, if I'm trying to troubleshoot something or if I captured something I was interested in it, that was always my go-to. Then, we have-
- Jon Krohn: 00:21:46 When would you do that? When in the process? Is this during the brewing process, would you just, if something looks unusual, something looks like it's awry or tastes like it's awry, or smells like it's awry? Or are brewers always doing these kinds of tests, plating yeast and bacteria, I guess? When does that happen in a beer brewing process?
- Beau Warren: 00:22:11 Well, I think ideally at every single step. Going into the fermentor, you want to make sure that wort is sterile, and you have a certain CFU, cell forming unit, on your plate.

Jon Krohn:	00:22:26	What's the wort?
Beau Warren:	00:22:29	Sorry. Wort is sugar water and hop. It's sugar from the grain, any hops if you use ... All beer pretty much has hops in it, so hops. Unfermented beer, I want to say.
Jon Krohn:	00:22:51	Got it, got it.
Beau Warren:	00:22:52	It's like sugar water, essentially.
Jon Krohn:	00:22:54	It's spelled W-O-R-T, if I remember correctly.
Beau Warren:	00:22:57	Yeah, that's correct.
Jon Krohn:	00:22:59	This is what you start with. You, as a human, can mix together your first set of ingredients in water. Yeah. The key ingredients in beer, I guess, are barley, malts, hops, and water, right? That's typically what you would put in, those four things?
Beau Warren:	00:23:16	Yeah. Then, people can throw in some adjuncts. Wheat, oats are used a lot in New England IPAs or the hazy IPAs. To increase mouth feel, a lot of people use rye or dab a little bit of a spicy character, not necessarily in New England IPAs. You see it a lot in stouts and stuff. You also see a nice roast, roasted barley for stouts. Guinness is well known for using roasted barley. You'll see some other kilned malts, dark in the color, but they'll caramelize it so it won't add a roasty bitterness, it'll actually add caramel. Or, caramel for some people.
Jon Krohn:	00:24:04	Okay, okay, okay.
Beau Warren:	00:24:06	Yeah.
Jon Krohn:	00:24:06	All these different things are added together into water, it's a sugary, sweet water. That's called the wort. That's, I guess, the starting point for brewing. Then, you add yeast

to it, to start fermenting the sugar water into alcohol and carbon dioxide?

- Beau Warren: 00:24:23 Yeah. We start on the hot side. So as you mentioned make that sugar water, boil it for an hour. Some people will do three hours, if you're doing some super stout or barley wine, or longer. Then, you throw it through a heat exchanger. Immediately cool it down to fermenting temperature, infuse it with oxygen for the yeast and then pitch the yeast into the wart that will be in the sanitary tank or the fermentor.
- Jon Krohn: 00:25:01 Cool. And yeah, every step of the way we could be doing like QAQC, looking at how the yeast is doing, I guess, and what kinds of bacteria are present. That isn't something-
- Beau Warren: 00:25:11 Yeah.
- Jon Krohn: 00:25:11 ... I'm really aware of. So both of these things are very small organisms.
- Beau Warren: 00:25:15 Yeah.
- Jon Krohn: 00:25:17 And yeast, I know, is the animal that is actively, because if I'm remembering correctly, yeast is like the simplest, it's one of the simplest animals. It's a single celled animal. So it's not a plant, it's not a bacteria, those are different kingdoms of-
- Beau Warren: 00:25:33 Yep.
- Jon Krohn: 00:25:35 ... animals or species. And so this tiny single-celled animal, the yeast, is, yeah, adept at converting sugar into alcohol and into carbon dioxide, that's how it survives. But why are there bacteria present? It sounds like that's a pretty normal thing. To me, that sounds like potentially an alarming thing, like you don't want a bunch of bacteria growth in your food typically.

Beau Warren: 00:26:03 Yeah, you definitely don't. And our processes at Aslim were very good, so we never really saw a ton.

Jon Krohn: 00:26:16 I gotcha, I gotcha.

Beau Warren: 00:26:16 And if it was there, it was under threshold, which is barely any. But yeah, if it's there, most of the time you don't want it there, some of it, it can't tolerate alcohol or hops.

Jon Krohn: 00:26:28 Right.

Beau Warren: 00:26:28 So what'll happen is you'll see it pop up on our plate and it's not going to be able to grow in beer, so that'll happen sometimes.

Jon Krohn: 00:26:39 Gotcha, gotcha, gotcha. So like-

Beau Warren: 00:26:40 Now, that doesn't mean that like everything's perfect and dandy, you want to get rid of it, but there's other things like lactic acid producing bacteria that will sour your beer and that is something that is like, giant red flag, such as *Pediococcus*, which is super, super, super small cell compared to a yeast cell. So that stuff, as you can imagine, can just embed itself into everything in the brewing process. And that can cripple a brewery eventually.

Jon Krohn: 00:27:14 Wow. Because it's just kind of-

Beau Warren: 00:27:14 Yeah.

Jon Krohn: 00:27:14 ... it becomes like a pandemic, like a bacteria pandemic in the brewery.

Beau Warren: 00:27:19 Yeah. And I mean, unless like-

Jon Krohn: 00:27:20 Wow.

- Beau Warren: 00:27:22 Yeah, I mean, like most places luckily have checks in place and are able to detect it and narrow down where it's coming from. But yeah, there's a lot of stuff going on in beer.
- Jon Krohn: 00:27:34 Yeah, like living stuff, and so sometimes things can get out of control, I guess. So we've gotten this sense now that the quality control involves monitoring for bacterial growth. Are you also monitoring as part of your QC process, literally things like flavor? Like do you taste things along the way to make sure that things are kind of coming along or what other the kinds of quality control steps are there as you go through the process?
- Beau Warren: 00:27:59 Yeah, definitely. And as we're about to talk about this project, the most important thing is if it tastes good. Because-
- Jon Krohn: 00:28:09 Right.
- Beau Warren: 00:28:09 ... even if all your checks are perfect and you're hitting target and everything, nothing weird is grown on plates. Your dissolved oxygen levels are great at canning. If the beer tastes like crap, then it doesn't matter how it looks on paper.
- Jon Krohn: 00:28:30 It's completely bacteria-free. Enjoy your beer. Enjoy. Who cares about the flavor? There's no bacteria. Okay, cool. All right, so I'm starting to understand this a bit better, starting to piece things together here hopefully for our listeners as well. So when you're brewing beer, one of the key things that people hear about is ales versus lagers. I think the popular perception, people will say, "I like lagers," or, "I like ales." And I think that's because in the mass market, lagers are typically relatively light-colored beers. The one that I had, but have now finished, it was a lager, so it was that light-colored, people can go back to

the beginning of the episode to see when I had a full glass.

00:29:19 And then ales are often on grocery store shelves and beer store shelves like a darker color, and maybe they have a bit of a richer flavor, a sweeter flavor. So I think people think in their heads that that's what the difference is. But my understanding, and I've never brewed a beer in my life, but lager versus ale, it's really just to do with one of them is top fermenting. The other one's bottom fermenting. And I can never remember which is which. But you could actually, and correct me if I'm wrong, but can you basically get any kind of flavor out of a lager or an ale or these kinds of distinctions of lagers being light, ale's being dark, that's kind of like an, that's an oversimplification, right?

Beau Warren: 00:29:59 I think it depends. So your macro-lager, super clean, super light, it's pretty much got zero sugar left because the yeast have pretty much-eaten everything because they added in, it let's these places, these macro lagers, which we make these style lagers too, based on one of these models I've made, but I'll talk about that later. They're super clean, easy drinking, no phenolic off flavor, no weird esters or flavors, super clean. And the reason they're clean is because of *Saccharomyces Pastorianus*. And *Pastorianus* is very efficient at fermenting at cold temperatures. And the reason *Pastorianus* is very good at fermenting cold temperatures is because you have *Saccharomyces Eubayanus*. So *S-eubayanus*, and then you have *Saccharomyces Cerevisiae*, which is your ale strain. These two at some point, they made a hybrid, so they mated and it must've been a really harsh environment. And then it's formulated and they created this new offspring called *Saccharomyces Pastorianus*.

00:31:27 So *Pastorianus* that everyone uses is actually a hybrid strain. So what it did is it got its cold fermenting tolerance

from Eubayanus parent, and then it got its maltotriose complex sugar ability to metabolize from its Cerevisiae parent. And you combine those two and you have a very efficient machine at low temperatures. And because it's at lower temperatures, it's going to be a lot cleaner, not a ton of flavors and esters going on. I like to explain that the reason ales have more esters and flavors is it's kind of like an equivalent of if you're moving furniture and it's like 80 degrees or a hundred degrees in your house, you're going to be sweating a lot. And the sweat equivalent is the esters and flavors.

- Jon Krohn: 00:32:41 Eager to learn about Large Language Models and Generative A.I. but don't know where to start? Check out my comprehensive two-hour training, which is available in its entirety on YouTube; yep, that means not only is it totally free, but it's ad-free as well. It's a pure educational resource. In the training, we introduce deep learning transformer architectures and how these enable the extraordinary capabilities of state-of-the-art LLMs. And it isn't just theory. My hands-on code demos, which feature the Hugging Face and PyTorch Lightning Python libraries, guide you through the entire lifecycle of LLM development—from training to real-world deployment. Check out my “Generative AI with Large Language Models: Hands-On Training” today on YouTube. We’ve got a link for you in the show notes.
- 00:33:24 Gotcha, gotcha, gotcha. That's super interesting. Okay, so the lager yeast, which you pronounced it there, I'm not, I would just completely butcher it, but the lager yeast, say it again?
- Beau Warren: 00:33:38 Pastorianus.
- Jon Krohn: 00:33:41 Pastorianus.
- Beau Warren: 00:33:41 Yeah.

Show Notes: <http://www.superdatascience.com/755>

- Jon Krohn: 00:33:42 It is able to ferment at very cold temperatures, which using this analogy, and this is actually, that's going to help me now remember that it's a really clean favorite, you're so using Fahrenheit, which half of our listeners are in the US that'll make sense to half of our listeners. In Fahrenheit temperatures when it's like 40 degrees, 50 degrees, it's a pretty cool day that's about five degrees or 10 degrees Celsius. And at those kinds of temperatures, if you were to do a bunch of moving furniture around from one apartment to another or whatever, you wouldn't get very sweaty. And so that's kind of what's happening with these Pastorianus, this lager yeast. It's getting, at low temperatures, it's able to do a lot of work efficiently converting the sugars in the wort into alcohol and into carbon dioxide, giving beer its alcohol and fizziness, respectively.
- 00:34:46 And so yeah, so you're typically getting with lagers a nice clean flavor. So yeah, you're completely ruining my hypothesis around lagers and ales, not necessarily having different kinds of flavors. But then ale's on the other hand, it's like moving on a very hot day, moving that furniture on a hot day. So if you're talking about an 80 or a hundred degree Fahrenheit day, that's like a 30 degree Celsius day, very hot outside, you're getting really sweaty. So similarly, they eat yeast, they're just shedding all kinds of interesting flavors into the beer and giving ales, which I tend to prefer, their yummy flavors in my view.
- Beau Warren: 00:35:29 Now, think about this too. When you're in a hotter environment, you're able to move around a little quicker and get stuff done faster.
- Jon Krohn: 00:35:38 The joints are all lubricated.
- Beau Warren: 00:35:41 So lagers are a little slower in most cases than ales can be. And then you also have to think about, okay, just like humans, some humans are just predisposed to sweat

more, even if it's like one degree outside. You know what I mean? So some of these other strains, these cold-tolerant strains, will still impart estrus and flavors reminiscent of an ale strain. So Eubayanus, which is the parent of Pastorianus, tends to have, they call it phenolic off flavor, just because it naturally expresses it due to its nature. That's where it gets kind of confusing. I'm not going to dive deep into this right now, but yeah, so in a general idea, that's the best metaphor I can really figure out for the understanding of this.

- Jon Krohn: 00:36:50 Yeah, no, that makes sense. I've definitely learned a bunch there. I've been misinforming people, lagers and ales. But one of them is top fermenting and the other is bottom fermenting, right?
- Beau Warren: 00:37:03 Yeah, I'm trying to think. The only reason one is considered top fermenting is because there's a froth in, there's a really thick froth in an ales because all the cells tend to, they're flying all over the place and it tends to foam up at the top.
- Jon Krohn: 00:37:22 Frothing is like a foamy layer on the top of the-
- Beau Warren: 00:37:25 Exactly. And that's mostly cells up there, which is true. But lager strains, they don't just sit on the bottom.
- Jon Krohn: 00:37:36 Oh.
- Beau Warren: 00:37:36 They're all over the place, but they don't make a big thick froth like ales do.
- Jon Krohn: 00:37:41 I see, I see. So it's a visual thing, so-
- Beau Warren: 00:37:45 Exactly. Yeah.
- Jon Krohn: 00:37:45 People call ales top-fermenting because you're seeing all this crap on the top, all these cells bouncing around. Gotcha, gotcha, gotcha. But it's an oversimplification to

call ales top brewing and lagers bottoming brewing. Cool. All right, so that is super helpful for me. So now let's start to get into this machine learning project. So at a high level, the idea that you had, so from your experience brewing at, prior to Species X, you have a data set of several hundred rows where every row represents a different beer where that beer went out to the public and the public tasted it. So for each of those hundreds of rows, you have a couple hundred, well, I don't know, maybe not a couple hundred, several dozen columns of information.

00:38:46 And those things relate to things like what was the alcohol level, which is kind of easy to understand. So I'm looking at one of these spreadsheets right now, and some beers have as little as 2%, some have 7%, some have 10%, but there's lots of other features in the dataset related to ingredients that went in, things that relate to the process that was followed, like temperatures that were used, ratios of different things to each other in the brewing process, so it provides this really rich feature set for training a machine learning model. And then Beau, the really brilliant insight that you had was taking all of these features, which is already a rich proprietary dataset. And the way that it's so well organized is mind-blowing to me. It must have a huge amount of work, went into organizing all these inputs into a machine learning model.

00:39:44 But then in order to be able to use all of those inputs, you could use an unsupervised machine learning approach that doesn't have labels, and you could end up doing an exploratory analysis that would group related kinds of beers together based on these feature sets. But ideally, for brewing the most delicious beer possible, which is probably what a brewer is typically trying to do, it'd be great to have a label as to how delicious this beer is, and you obtained that. So you have this outcome, this label

for all of these hundreds of rows that's based on consumer feedback. So it's a score, a score rating, and it's a proprietary nature to this, but so I'm not going to go into too much detail of the scale, but it's a numeric score that is consistent across all of these beers. And so it allows you to have a rating of what people think of the beer, how delicious it is.

00:40:51 And so for these several hundred rows, all representing beers, you've got tons of rich features going in, and you've got this outcome that you can predict with a supervised learning model. And so you blended these model inputs with this feature set. And yeah, you've been training a machine learning model. And one thing that I want to, so I'd love for you to dig a bit more into some of the key features. Obviously, we can't go into all of the dozens, and to some extent you might want to keep some of it proprietary, but giving us a sense of what some of the key features are in the inputs to this machine learning model. But then the second thing is that I'd love for you after you've explained the features to talk about the machine learning process. And something that's really amazing for me is that we haven't talked about this since you and I have been in the conversation. I would've mentioned it quickly in the episode's intro, but you don't come from a traditional kind of machine learning background.

00:41:56 You weren't a statistician before, you weren't a computer scientist, before you were a football player. You had an opportunity to go pro and play for the Rams in the NFL. You come from a family of football players. So you played college football with your two brothers. Your dad was a professional football player for the Washington at the time, what they're called the Redskins football club. And so yes, you come from this football background, you studied psychology and sociology in college while you're playing football, and then you became a brewer because of your passion for brewing. But I guess in recent years,



because of your love of science, you became interested in AI. And so you have been self-studying machine learning, and you have obtained some credentials, but not getting a degree. So you have a certificate in the practice of data and analytics from Ohio State University, which is cool. And I understand that you have more ambitions to study more.

00:43:02 But the point that I'm getting to here, and I've been talking for a very long time, but the point that I'm getting to is, and I hope this is inspiring for any of our listeners out there, if you are interested in using machine learning in some application area. So in Beau's case, he happens to be passionate about beer. And so he was like, how can I use machine learning and AI to brew better beer? And he went out, he started listening to this podcast, to The SuperDataScience podcast. He started studying things online and having reviewed your code and gone through your process with you, I can say it's amazing. It is as good as the professional data science work that I see on a day-to-day basis. And well, I can't wait to find out what the results are like. But yeah. So I've now talked for a very, very long time. I don't know if there's anything else about your background that you want to talk about, but your background is something we'd love to hear about the features that are going to this machine learning model, and then of course talk about the machine learning model itself.

Beau Warren: 00:44:05 Okay. Yeah. So when I first came up with this idea, I was like, man, this would be fricking awesome. I think this was when we started seeing Dolly come out and I was like, I just need to start this project up. So I started gathering all the proprietary data I have from brewing recipes for the past, I think I was doing it for about seven years then, so it took me six months to clean all the data.

- Jon Krohn: 00:44:42 Wow. I'm not surprised. That's something that's one of the things I didn't know that, but it's one of the things that blew my mind about this whole project was I was like, it's mind-blowing to me that somebody had the foresight to be putting all these data together over the years in this structured way, but that wasn't the case. It was retrospective mostly. So I imagine now you do a better job of like, okay, we brewed a new beer, let's get it into the right format into this dataset.
- Beau Warren: 00:45:06 Exactly. Yeah. So that's actually why it's funny because it's actually pretty similar to brewing in that it's like 80, 90% cleaning, and then 10% is actually executing, actually brewing, which is definitely true when it comes to this. That's where these two things definitely align.
- Jon Krohn: 00:45:28 So when you're making beer, this isn't something I was aware of. So our listeners who are data science practitioners will know that a huge amount of being a data scientist is cleaning the data and actually training the models is typically a small part of the job. So when you say cleaning in a brewery, you literally mean like, okay, we finished brewing this beer. You just need to spend tons and tons of time cleaning everything and getting it back to spic and span before you start over.
- Beau Warren: 00:45:52 Yeah, I mean, the tanks and stuff, the ground, the outside of the tanks, all those. Now granted in a professional brewery it's a little easier, but still it takes a lot of cleaning and maintenance. So yeah, a majority of it is cleaning, making sure everything's sanitary, that sort of thing. So yeah, when I came up with the idea, it had to work with, so when I was setting up the data, the feature columns, there actually is 180 columns total.
- Jon Krohn: 00:46:33 I had this idea in my head that it was like 200, but when I just was skimming over now.

- Beau Warren: 00:46:37 Yeah, it's about 200. But what I'll do is when we put guardrails up, depending on which model we're using, we'll shrink the data, the training data, or just keep it the same. If we're using Behemoth, which is the model with every single bit of our data, or if we're doing Chr15 P is my lager model, Chr15 P, will narrow down the rows, so it just includes lagers. And then will shuffle around and delete out some of the feature columns so that it's all related to lagers only. Where Behemoth is just like, okay, how do I get from A to B to the most customer satisfaction with no disregard to style or anything. So that's what Behemoth is. And usually what comes out the other end is 10 to 15% alcohol beers, and a lot of stout-influenced beers, but most of them are pretty off the wall crazy. And we'll try to navigate it to the ones that are practical to brew. So the one from Behemoth, for instance, that we'll be brewing here for the first time is I think it's a 10, 8.5% Baltic Porter, which is a lager, lager strain. It's got 20 different malts in it. There's dashes of certain malts, which I find fascinating. There's like lactose, marshmallows, vanilla, three different types of fruit. It's got a big whirlpool charge, actually just a normal whirlpool charge of hops, which is really fascinating.
- Jon Krohn: 00:48:39 What does that mean, whirlpool charge? Everything else kind of made sense to me but-
- Beau Warren: 00:48:43 The whirlpool charge is just like a dosing of hops in the whirlpool stage of the beer.
- Jon Krohn: 00:48:48 Yeah. And what's that?
- Beau Warren: 00:48:51 So we'll hook it up in the kettle and we'll circle it after it's done boiling. And at that point you can actually cool the wort down. So we'll pull it down to a certain temperature and then add hops if we need to. Usually whirlpool charges happen for IPAs only. So this is why it was fascinating to me that a model like Behemoth throws a

completely novel way of brewing a Baltic Porter, especially at me to brew. And this is where I think this is really novel stuff. Some of these styles, they're not anything I've ever seen before. And as a domain expert, I guess you could call me just from brewing for several years or seven, eight years now, I can look at a recipe that it generates and be like, wow, actually, this is doable and it sounds pretty nuts, but I want to brew it and see how it goes. So there's that generative, I don't know, I would call it a generative AI, but it's definitely supervised. Definitely supervised machine learning.

Jon Krohn: 00:50:14 Yeah. Typically, we would reserve generative AI for something that's generating like tokens, like natural language or code or something. It's typically what that means, but it is generative in the broader sense, in the sense of it is actually it is helping you generate new beer recipes. And that's a key part of your machine learning process here is that you have a simulation stage, and I guess maybe you could kind of consider that generative. So there's a process in your algorithm where you'll generate say, 10,000 or a multiple of 10,000 random rows that are permutations of all of the relevant features. So I guess for Behemoth, you have all 180 features going to the model. And so your simulation stage is creating 10,000 or more random permutations.

00:51:17 But another really clever thing you did, and this leverages your domain expertise, is you set reasonable boundaries on those simulations. So in that sense, it's kind of like what a Bayesian statistician does. So a Bayesian statistician will have these priors around parameters or in their model that kind of constrain how far a parameter can go away from, so you could have a distribution that describes the probability of where a parameter could be. And so that's similar to what you've done here with your simulation. So I'm kind of jumping the gun here, but you have the simulation stage, and then later you're able to

use the supervised model to say, "Okay, with these 10,000 simulated recipes." Based on the constraints that you have, what is the likely rating that a user would have based on these inputs? I don't know if I cut you off or if you want to get into a bit of what that supervised learning model is like because I've tried a bunch of different approaches. I think it would be kind of cool for us to... If I haven't interrupted you too much and I'm not changing the topic too much, it'd be cool to go into the specific libraries that you used in order to create your supervised learning model.

- Beau Warren: 00:52:40 Yeah, definitely. So Darwin, as a name suggests, that's our fruited sour model. Darwin is completely, it's all genetic algorithm, so it only uses genetic algorithm. So TPOT is a really good way to utilize genetic algorithm. I'm assuming you might have come across genetic algorithms before?
- Jon Krohn: 00:53:12 Yeah, I mean, I can talk about it a bit if you like. So the idea with genetic algorithms is that you start with a random starting point with a bunch of different models, and then you see how they perform. So you have this random starting point for a bunch of different models and you see how they perform at the task that you'd like them to. And you take the top performers similar to a way you might, if you're trying to make a fast racehorse or you're trying to make a yeast that ferments at a really low temperature, you'll take from all of the possible species, you can call these random starting point algorithm species, and you take the ones that are great performers. So just like the horses that are fastest or the yeast that bruise at the coldest temperatures, you'll take, in this case, the models that are best at accurately predicting how people will rate a beer.
- 00:54:21 And so that's like your first generation. You've now taken these top performers and you breed them together. That's

the key thing. These genetic algorithms is you then say, "Okay, let's take two top performing models from my first generation." Blend them together in some way. So there's some attributes from one parent and some other attributes of the model from some other parent, and you blend them together to get the second generation. And some of those second generation ones, you're going to end up having blended the wrong parts of the parents together by chance, because it all happens by random assortative mating, just like in the real world where chromosomes are sliced and diced at random to create a child. And so you're going to get in that second generation, some that perform about as well as the parents did in the first generation. Some are just going to completely bomb because they got the wrong parts of the model process from their parents.

00:55:20 But some of them are going to outperform because they happened to get the really good stuff that their parent one had and the really good stuff the parent two had and they blend together. And so in that the second generation is going to result in some child models that are better performers than anyone in the second generation. And then we just repeat this process over and over. So you take just like breeding horses or making big apples or having cold fermenting yeast, you're picking the ones that are the best from each generation, mating them together. Yeah. So that's cool. So you call your genetic algorithm, Darwin. And the key thing here, I think one of the key takeaways other than in general, how evolutionary algorithms work is your recommendation to use the TPOT library, which I presume is a Python library.

Beau Warren: 00:56:11 Yeah, it's a Python library. It pretty much sets it up for you, just lays it up. So all you really have to do is define the TPOT algorithm, your training data, and then your parameters are really easy to adjust on the fly. So if you want to mess around with your offspring or a number of, I

think it's mutation rate, you can definitely do that. And I've found for TPOT specifically, that genetic algorithm, it has a really solid default set of parameters that we're good out the gate. So if people are interested in genetic algorithms and they don't have a ton of experience like I do, it's definitely cool to learn from and utilize TPOT.

- Jon Krohn: 00:57:11 Nice. Very cool. Yeah, I've actually never used an evolutionary algorithm on a project, so it's cool that you got to do that. And then also, I think you made extensive use of the PyCaret library, if I remember correctly.
- Beau Warren: 00:57:24 Yeah, I use PyCaret, I use H2O AutoML as well. This is just for exploratory. So I use PyCaret and some other AutoML just to find some base models. So just to get an understanding of what works best on this training data, I mean, most of them seem to be... Actually a lot of them, I've been getting boosted trees, so I'll take a look at that. And for instance, if it's a boosted tree that comes out on the other end as the best score, I'll take that and then run a separate model on that specifically to do a grid search and find the optimal parameters that way as well. So, I'll brute force it.
- Jon Krohn: 00:58:31 Fantastic. So to break this down a little bit, so you use tools, AutoML tools like PyCaret and H2O, both of which I will put in the show notes to allow you to identify models that might be useful for say whether it's Behemoth or Chr15 P. So Behemoth working across all of your features, or Chr15 P working with a subset that are relevant only to lagers. And also the dataset size is being quite different there. So Behemoth, there's a lot more rows to work with, whereas Chr15 P is only working with a lager data, which turned out to be a relatively small proportion. And something that I was like, "I don't really like IPAs." And you're like, "Well dude, almost all of the beers that we have are IPAs." So we're going to have to reduce down the dataset quite a bit. And so you end up in

those different scenarios, different kinds of models are going to end up being great, but we don't know typically as the human in this particular scenario, if I've got the larger situation with a much smaller number of rows of data and a somewhat smaller number of columns of data for this particular circumstance, does that mean I should be using the same model as I was for Behemoth? And maybe you could spend a lot of time messing around with different kinds of models to figure out what the best approach is for Behemoth in the first place.

00:59:53 So to sidestep all of that, you quite cleverly used AutoML packages like PyCaret and H2O, in order to be able to home in on what a great modeling approach would be. And as you're saying there, boosted trees ended up often being the approach, which isn't surprising. People who know from data science competitions like Kaggle, these kinds of boosted tree approaches like XGBoost end up often being top performers. And it looks like from your code that you're typically using Scikit-learn as the library for finding a tree's module.

Beau Warren: 01:00:43 Yeah, yeah. Definitely use Scikit-learn quite a bit. I don't know. Every now and then we'll come across some random one that scores really well despite using H2O and PyCaret, but will both score some off the wall algorithm that I haven't seen before or heard of before, like OMP. It's a greedy algorithm. I forget exactly what it stands for, but I thought it was a fascinating algorithm. But every now and then, yeah, you get some wild algorithms. Also, what I do sometimes is, okay, so I'll do the AutoML run. I'd be like, okay, for instance, XGBoost did great on this, so we'll move it to the brute force method and use grid search, and then I'll experiment and ensemble that with something else. So one that I've really enjoyed using a lot is TabPFN, which is a neural network, which is pre-trained on tabular data. If I'm correct, I might be [inaudible 01:01:56] that wrong but-

Jon Krohn: 01:01:57 Yeah, yeah, yeah. You're correct. Absolutely. Yeah. I looked into this when you first mentioned it to me because I had not heard of TabPFN before. So typically neural networks are not good at working with tabular data. And in fact, up until you mentioned this, I was not aware of people using neural networks, the Deep Learning for tackling tabular data problems. I wrote the book Deep Learning Illustrated, and I've taught Deep Learning tons of times. And I had students who regularly, as I was teaching Deep Learning courses, intro to Deep Learning courses, they would be interested in tabular applications. And I'd be like, "Yeah, yeah, yeah, yeah, yeah, yeah. Maybe [inaudible 01:02:35]."

01:02:36 We give it a shot, but it's probably not going to be very fruitful. You're probably better off using decision trees, for example, a boosted tree approach like XGBoost instead. But yeah, so Deep Learning is usually better for pattern recognition where there's a lot of structure in the data patterns that can be learned spatially in a way. So it's easiest to understand with machine vision problems. So when your inputs are images or video, there's a lot of spatial patterns that Deep Learning algorithms are adept at being able to tease out what the most important signal is from amongst all of that noise going into the model. And a similar kind of thing ends up happening with natural language processing, which Deep Learning has also proved, as we can see with the large language models that use Deep Learning under the hood. Well, there are, it's not even under hood. Fundamentally, large language models are Deep Learning networks. With natural language processing, with machine vision, we see amazing powerful real-world applications in machine vision and in generative AI, for example. But tabular data, where the rows are typically, they're not necessarily related to each other in the way that the pixels in an image are or the words in natural language are. And so yeah, up until you mentioned TabPFN to me, I wasn't

aware that somebody had come up with a way of doing neural networks effectively on tabular data. So it's really cool that you found that. I'm really glad you brought it up. And yeah, you probably have more to add.

- Beau Warren: 01:04:20 Yeah, I love TabPFN. It's limited on the number of feature columns you have, as well as rows. So is my data, you know what I mean? My training data is not massive, so it is actually great at ensembling with other algorithms either to kind of hone them in to be more accurate, I have found. So that as an ensemble has been pretty eye-opening, and the fact that it's instant, TabPFN literally takes half a second to train and to make an inference. It's super fast, mind-blowingly fast.
- Jon Krohn: 01:05:19 Yeah, I guess with the relatively small amount of data points that you have, it's not like with the other approaches when you're talking about doing the boosted trees and particularly using the GridSearchCV method, the model selection module of Scikit-learn. When you do a kind of grid search like that, you're going to be running so many different models, it's going to take a long time. The cool thing about Deep Learning, and part of what makes it so powerful is that it allows for features to interact with each other in all kinds of crazy nonlinear ways. And so it's kind of like you're running lots of experiments, especially how non-linearities and interactions relate to your outcome in a Deep Learning model. So it's cool to hear that it's so efficient on the relatively small amount of data in TabPFN, and it doesn't surprise me that that's the case.
- Beau Warren: 01:06:14 Yep.
- Jon Krohn: 01:06:15 Very cool. Yeah, so I hope that that gives our listeners, obviously it's not like we can share Jupyter Notebooks with people unfortunately, or go into too much more detail on what you're doing because otherwise you'd be

giving away the secret sauce. But some really great pointers here, general machine learning ideas, TPOT for evolutionary models, PyCaret and H2O for AutoML, Scikit-learn for a lot of your machine learning implementations, including, as you mentioned, boosted trees. I know also that sometimes K-Nearest Neighbors algorithms did come up as a top model in some of your situations. And TabPFN, super cool recommendation for fitting Deep Learning models to tabular data. So I'll be sure to include links to all of those Python software libraries in the show notes and all them are open source, so you can pick them up right today and start playing around with them.

- Beau Warren: 01:07:14 So another one is called Lazy Predict. It's an amazing Python library. And this goes along the same tillage of PyCaret and H2O in order to find a high-scoring base algorithm to work with. And it'll list them out and show you the top scores. And usually what I'll do is I'll run two or three of these AutoML algorithm search utilities or libraries, and I'll usually run those side by side to see, okay, this is actually legit. I should be doing XGBoost, or I should be doing ADA or linear regression, even though I haven't seen linear regression pop up ever, but who knows? So anyways.
- Jon Krohn: 01:08:17 Very cool. And so you and I have had a few meetings, and so we were like, "Okay, let's brew." I agreed to help out and give you a bit of feedback on what you've been doing, the great impressive work you've been doing on your data collection and your modeling. And so we were like, "Okay, so we're going to have this collaboration beer." And as I already mentioned, I was like, "I don't really like IPAs typically, so let's remove them from the data." And so you've been working on a lager. So we reduced the data down to just the lagers that were in the dataset and you trained, so you used these AutoML processes to identify Chr15 P, your lager-only model, supervised learning

Show Notes: <http://www.superdatascience.com/755>

model for predicting how people would probably taste one of your tens of thousands of simulated recipes. So how's that coming along? Where are you at the time of recording? Where is the beer? And I guess we can tell people the name of the beer.

- Beau Warren: 01:09:37 Yeah, I think so. So I tasted this beer yesterday for the first time, and out of all the lagers I've ever made, it's up there. It might be the best lager I've ever made.
- Jon Krohn: 01:09:54 That's unreal. Wow. I'm so happy to hear it. I was worried that we'd get on air, do all this rigamarole, then you'd be like, "Ah, we got to go back to the drawing board. We should have been using K-Nearest Neighbors instead of boosted trees after all. Why did I ensemble TabPFN?" We can talk about this in a bit more detail. When you and I were looking over the results that came out of your proprietary supervised learning process. So you had these tens of thousands of rows simulated using the parameter constraints that you had on inputs, and then you had your Chr15 P supervised learning model go over all of those 10,000 simulated rows and predict what people would rate. One of those came out by some margin at the very top, and that's this beer that you've been brewing. But there were some things about the features in it that to me, it doesn't mean anything.
- 01:10:58 When I'm looking at the inputs into your model, most of them don't make any sense to me. I'm like, "Okay, I get the alcohol content." That one makes sense to me. And through going over your sheet a few times I've learned, okay, these are the hops and this is the massive hops, but that doesn't mean anything to me. I can't infer how that could possibly relate to the way that a beer tastes. But so as you were reviewing this lager that came out number one, by far in terms of how consumers are predicted to rate it. You saw some unusual things, your

eyes kind of lit up. What are those kinds of things that you saw?

- Beau Warren: 01:11:39 Well, the main thing was Marris Otter it called for, which is really fascinating because a majority of lagers do not use Marris Otter that's mostly bound to stouts and pale ales and IPAs. So I thought that was really interesting. Not too crazy, but I thought that was fascinating. There was, again, a really large hop edition, large for a lager, a large hop edition in the whirlpool, not a ton of hops in the kettle and no dry hopping. And in addition to that, there was also, it called for sugar as well, a very certain type of sugar, which was called a D-45 Candi Syrup. So we threw that in as well. And I forget what else, I'll have to take a look at that later. But I think there's different types of malt in there. Nothing too crazy, but I think there was wheat in there as well. So anyways, as I'm brewing this and throwing Candi Syrup in it and doing the weird whirlpool charge, I was like, this is so bizarre, but in a really cool way. I feel like there's an alien telling me what to do, and it makes sense. I just know that a person would never suggest brewing this usually, but it's coming out amazing. It's so unbelievably smooth, really light, and has a nice depth of character and flavor. And it's just, as I said, super clean, easy to drink. So we'll go ahead. It's about done fermentation, so it'll be conditioning for a couple of weeks probably before we tap it and package it. So-
- Jon Krohn: 01:14:10 Amazing.
- Beau Warren: 01:14:12 ... I'm super excited about it.
- Jon Krohn: 01:14:14 I've had an ear-to-ear smile as you've been describing this. It's been such a wild process. It's quite different. It's completely different from any kind of project that I've ever been associated with. And I played an extremely small part in this. This is really your project. You've done all the

hard work on this and it's fantastic to see what's come out. But that particular thing that you're describing there, and I think that shows the real value in what you're doing at Species X, at least on the silicon side of things with the silicon species. It's just like you're saying like an alien directing you what to do. And so you've spent years now, particularly quite a bit of time over the last year, though. That doesn't include all the time you spent getting the data together, but you've invested years in aggregate into data collection, modeling, and also just training yourself on the side. And for that to come together and produce this result, which I guess we still haven't had the final taste, but it's come along now that you have a pretty good sense of it.

01:15:27 For me, the two most exciting points in this whole process were going over this top-rated result that came out of your Chr15 P model. The beer that had selected seen your eyes light up as you talked about things like the Maris Otter, which is a particular kind of barley malt. And as you say, a flavor that people wouldn't think to typically put in a lager combined with things like this specific syrup that you mentioned. Lots of hops added at the whirlpool stage, which I was like, "Oh, man, I don't like IPAs. And now, the algorithm has suggested all these hops that ruined everything after all." And you're like, "No, it's okay. Because at the whirlpool stage, lots of hops, they won't create too much bitterness in the flavor like they would if they were added earlier on in the process." And so I was like, "Okay, cool. Let's go for it. This sounds like a beer based on what you're describing that I would like."

01:16:18 And so that was one really exciting stage in this. And then the second really exciting stage is right now. To be able to hear that it's working out because you don't know for sure with machine learning models. And most historically, for me, in fact, maybe 100% of the time with

any other machine learning model I've ever been involved with in more than 15 years of working on commercial machine learning projects. All the other times I can know what the results are going to be like right away. I can see them on my screen. Does this work or does it not work? But with what you've done, it was going to be a couple of months from like, "Okay, this is the recipe it suggested." It's like a marionette to Beau, like an alien telling you what to do as you go through creating this beer. And you don't really know how it's going to turn out until the end. And so for it to be seeming like it's going to work out is super exciting. And so it sounds like February 8th in Columbus, Ohio at the Species X Brewery. If we have listeners in Columbus, Ohio, actually I do know that we have some Columbus, Ohio listeners. And I'll be reaching out to friends that I have in the city. And some friends may even come from far and wide of mine to come on February 8th and try out this Maris Otter lager that your Chr15 P algorithm has suggested. Can we tell people the expected name of the beer?

- Beau Warren: 01:17:52 Oh, yeah. It's called Krohn&Borg if you want to go into depth on that.
- Jon Krohn: 01:17:58 Yeah. So this is actually hat tip to my sister Stevie for coming up with this idea. So there is of course a very famous French beer called Kronenbourg out of a city called Kronenburg. And so it's a very popular lager, particularly in Europe. But you see it on a lot of taps in the US as well. It's a pretty common like you'll see it alongside Stella's and Heineken's. So yeah, get your Kronenbourg beer. And then, so the idea here is it takes advantage of my last name, Krohn. It helps people nudge them in the direction of pronouncing it properly, which is nice for me. And then it also adds in, if people are familiar with Star Trek, the Borg are these robots that become a big problem in the Star Trek series. Specifically, the one that I grew up on Star Trek: The Next Generation in the

'90s. They're like the arch nemesis to Jean-Luc Picard, and the Star Trek and the robots. So it's this idea of AI blended with humans Krohn and Borg and kind of rolls off the tongue. I wanted to call it Krohn&Borg 2064 to play on. They call it the Kronenbourg's 1664 for their lager, but your marketing team said that that might be cutting it too close.

- Beau Warren: 01:19:20 Yeah, it might be a little too close. Yep.
- Jon Krohn: 01:19:23 Treading on trademark issues, but that would have been a funny thing. But yeah, Krohn&Borg. I'm so excited, Beau, for February 8th to come out to Columbus and try it out. I'll be recording a shorter episode, probably a Friday episode while I'm there getting some feedback on the beer and answering maybe people's data science questions together to round out. So there will be an episode coming in the future on how the Krohn&Borg beer turned out from this amazing project that I am so grateful you invited me to play a small part in. Thanks a lot, Beau.
- Beau Warren: 01:20:12 Yeah, I appreciate you to take the time to collaborate on this. It means a lot. And like you mentioned, I'm mostly self-taught in this. I have a little bit of formal education from OCU, but nothing super complex and over the top. So having you to come in and validate that what I was doing was the correct way to do things and making amazing suggestions into possible future iterations was really helpful. So I'll always be grateful for your time.
- Jon Krohn: 01:20:56 Yeah, my pleasure. And I'm definitely down for more of these. It's very exciting for me. I mean, I'm passionate about beer. I love it very much. There are a few things I like more in life than beer. It's a funny thing. People will be getting really fancy bottles of wine or whatever at dinner. And I'm like, "Just pour me a little taste because you're wasting that expensive wine on me. Just get me a

beer." No, that's what I am. I can be a bit of a beer snob. But yeah, super, super honored to be part of this, Beau. And I'm glad that it seems to be working out so far. Up until today, I didn't know how this was actually going to turn out potentially in terms of flavor, we still don't know for sure.

01:21:40 But I was just like, I didn't want to get my hopes up. I knew that there was a chance. This is an experiment. And so, I guess I still know that there's a chance that it could fall down, but it seems like things are coming together. And so I've been holding my breath to find this out. So it's a really exciting day for me. I'm going to be over the moon for a while, for sure. One last final thing before we start wrapping up this episode because you had an interesting insight for me before we started recording. And it shows how deep you get in the weeds on anything that you're thinking about is you had an interesting parallel between... We talked about how when you were a college football player, you were also studying psychology and sociology at Virginia Tech. And you made an interesting parallel before we started recording to the large language models like GPT-4 that we have today.

Beau Warren: 01:22:36 Yeah. So this actually relates back to the project that we're doing here too. I've been using an agent open-source code that will go out and do research utilizing Claude or GPT-4 or one of the big ones.

Jon Krohn: 01:22:57 That's an open-source agent?

Beau Warren: 01:22:59 Yeah, it's Baby AGI.

Jon Krohn: 01:23:02 Baby AGI, yeah.

Beau Warren: 01:23:04 So that's a good... It's iterate, it's like ripped off of it. So it's not exactly Baby AGI. It's my own repository that we use here in-house at Species X.



- Jon Krohn: 01:23:16 Cool.
- Beau Warren: 01:23:16 But I can't open-source it just because there's a lot of proprietary information inside of that repository. So yeah, it's definitely a rip-off of that. In addition to changes on the actual Baby AGI code, there's also a lot of prompt engineering. And I still can't believe I'm saying the word prompt engineering, but it's definitely a thing now. It's just crazy to me. So the goal of the agent is to go out. It's given a goal, it goes out, does research on the top beers. And then it'll create a brand-new recipe and then it will go ingredient by ingredient improving the recipe at each stage. So it'll keep going until you just tell it to stop. And then in addition to that, it'll tell me its name. And then it also prompts me with an image of itself to put into a image generator. So going off that, you'll see on the website it says, various agents. That's what we're talking about there. Well, eventually, I have to figure out how I'm going to release one of those agent beers. But in there, you can see even in the Baby AGI repository code, it'll say temperature. You can mess with the temperature in there. I've noticed when you up the temperature of the model, it tends to do some pretty crazy stuff.
- Jon Krohn: 01:25:07 So to dig into this, this is when you have a large language model, there's a key hyperparameters. So a key attribute that you can adjust called temperature. And I guess it actually, in a way, we can bring this analogy back to your sweaty yeast thing because it's like when you have cold water that you're brewing in, the reactions are slower, more predictable. And when you're doing the reactions like you might with a nail in a warmer temperature, the action proceeds more quickly, but it's more sporadic. So it's similar with temperature as well. So actually, if you turn the temperature on a large language model all the way down to zero, you will end up with a totally un stochastic model. It'll be a deterministic model where every time you run it, you'll get the exact same output.

Show Notes: <http://www.superdatascience.com/755>

01:26:07 So it's just predicting the most likely generative sequence, whether that's an image or text or code or whatever. You have your generative AI model doing whatever you have your large language model doing. And as you turn the temperature up, it will have more random outputs, so it will explore. So you'd end up with more creative images or more creative, more random stories. And yeah, so that relates, Beau, too.

Beau Warren: 01:26:40 Yeah. So when I was studying psychology, you talk about how as a person gets more... When their schizophrenia really starts to develop, you can actually see this in artists too that are developing schizophrenia. At first, their drawings will be very... They'll look very realistic or photorealistic in some cases. They look normal, essentially. But as the schizophrenia develops, you'll see a lot more colors and more subjective shapes and stuff. More abstract until eventually, it's completely unrecognizable and it is just all colors on a piece of paper. And no shapes, no nothing. So one example I saw was a person was drawing a cat and they were really good at drawing cats. And then eventually, the cat forms start to go away until it's just like blobs of color.

01:27:46 And going back to the LLM reference, I noticed when I up the temperature on it, it would talk to itself. So it would have conversations with itself, and none of it made any sense like a word salad is what people call it. So it would just chirp off and start making absolutely no sense whatsoever. I was like, "Gosh, this reminds me of schizophrenia. It's just strikingly similar to the development of schizophrenia." If you have a temperature on zero, like you were saying, very proper, you know what I mean? Will not do anything too crazy, but as you up the temperature as referencing back to the cat drawing. The cat drawings just got crazier and crazier and crazier until it's unrecognizable and makes no sense.

- Jon Krohn: 01:28:53 Yeah. So it's an interesting analogy. And yeah, I mean, at a surface level, we are certainly not psychiatrists or experts in schizophrenia in any way. But from the abnormal psychology courses that I took as an undergrad studying neuroscience, I can see the parallels that you're making. And maybe it's an interesting parallel. And I wonder if somehow there is something useful there for schizophrenia researchers, I don't know. There may or may not be. The way that we have animal models of diseases. So we'll use mice in particular, or sometimes rats as animal models of different diseases including psychiatric diseases. And so I wonder if someday we'll have machines that can substitute and maybe even be a better substitute because obviously mice don't have language, for example. But yeah, I wonder if there's people that they're working on these kinds of application areas where you can have an LLM that models schizophrenia. And somehow that provides some insight into the actual biological disease itself.
- 01:30:09 So yeah, interesting ideas there, Beau. Thank you so much for this fascinating episode. I hope our audience has enjoyed learning about beer and about how anybody can be training themselves to take advantage of machine learning and AI. Open-source tools that are out there, learning opportunities that are out there that are available for free online. And yeah, so you can leverage your domain expertise to create some really cool AI projects like you're doing at Species X. And yeah, hopefully we're going to find in a couple of weeks that Krohn&Borg is a delicious beer and that this project was a success. I'm super excited about it. Been a great day for me already hearing this positive news. Beau, as a regular listener to the program, you will be aware that I can't let you go without a book recommendation.
- Beau Warren: 01:31:02 Yes. So I love space too in addition to this. One of my favorite books I've read recently was Operation Hail Mary.

I don't want to give too much away, but it involves space, microbiology, and saving the world. And it's amazing. It's a fiction book, but yeah, it's one of my favorite books I've read in a while. I think they're making a movie with Ryan Gosling starring it.

Jon Krohn:	01:31:39	Oh, really?
Beau Warren:	01:31:40	So that's something else.
Jon Krohn:	01:31:40	Probably big.
Beau Warren:	01:31:41	That's something to keep your eyes out for.
Jon Krohn:	01:31:45	Awesome. All right. And Beau, so how should people follow you? Obviously, I'll include a link to speciesxbeer.com in the show notes. Where else should people connect with you?
Beau Warren:	01:31:56	I'm really active on Instagram, so @speciesxbeer. That's the same handle for all of our social media, so active on Instagram, very active on Threads. Here and there on Twitter, and then Facebook as well. But mostly Threads and Instagram, you'll be able to find us.
Jon Krohn:	01:32:22	Well, you get a prize for being the first guest to ever say that they actually use Threads.
Beau Warren:	01:32:28	Really?
Jon Krohn:	01:32:28	Yeah. I've had people, especially when Threads first came out, they're like, "I've created a Threads account." But yeah, you're the first one to actually see that they're using it, so that's cool. It's nice to know something [inaudible 01:32:37].
Beau Warren:	01:32:37	Now, I don't have a ton of followers on there. But yeah, I post there pretty frequently, and I don't know. I have

Show Notes: <http://www.superdatascience.com/755>



more followers on there than Twitter/X, so I'm just trying to...

- Jon Krohn: 01:32:58 Nice. All right. Well, all of our many Threads listeners to the podcast, they'll be checking that out right away. Beau, thank you so much for taking all the time today. And I'm looking forward to catching up with you in Columbus on February 8th.
- Beau Warren: 01:33:16 Thanks for having me, Jon. I appreciate it.
- Jon Krohn: 01:33:18 Well, I'm super, super excited to try the Krohn&Borg, Maris Otter lager in a few days. Wow. Unreal to hear that Beau's AI beer project is likely to work out on the very first attempt. In today's episode, Beau filled us in on how sweaty ales have more flavor and efficiency than cleaner fermenting lagers. How he painstakingly curated nearly 200 feature columns for 300 beers over six months. How he used the PyCaret, H2O, and Lazy Predict AutoML libraries, as well as the TPOT genetic programming library in Python for homing in on great models for his beer rating predictors. How he used scikit-learn extensively throughout his work, particularly the GridSearchCV method for finding optimal model hyperparameters. How he ensemble more traditional ML models like boosted trees with TabPFN deep learning models to get even better results. And how you can join us in person at Species X Brewing in Columbus, Ohio on Friday to taste the resulting Krohn&Borg beer that emerged from these years of AI, R&D effort from Beau.
- 01:34:25 As always, you can get all the show notes including the transcript for this episode, the video recording, any materials mentioned on the show. The URLs for Beau's social media profiles, as well as my own at superdatascience.com/755. Thanks to my colleagues at Nebula for supporting me while I create content like this SuperDataScience episode for you. And thanks of course

Show Notes: <http://www.superdatascience.com/755>



to Ivana, Mario, Natalie, Serg, Sylvia, Zara, and Kirill on the Super Data Science team for producing another delicious episode for us today. For enabling that super team to create this free podcast for you, we are of course, so deeply grateful to our sponsors. You can support this show by checking out our sponsors' links, which you can find in the show notes.

01:35:05 And if you yourself are interested in sponsoring an episode, you can get the details on how by making your way to jonkrohn.com/podcast. Otherwise, please do share, please review, please subscribe in all those good things that help us get the word out there about the show. But most importantly, just keep on tuning in. I'm so grateful to have you listening, and I hope I can continue to make episodes you love for years and years to come. Until next time, keep on rocking it out there. And I'm looking forward to enjoying another round of the SuperDataScience podcast with you very soon.