



# **SDS PODCAST**

## **EPISODE 922:**

### **AI FOR**

### **MANUFACTURING**

### **AND INDUSTRY, WITH**

### **HUGO**

### **DOZOIS-CAOUCETTE**



Jon Krohn:	00:00	Welcome to episode number 922 of the SuperDataScience Podcast. I'm your host, Jon Krohn. Today's guest is Hugo Dozois-Caouette, co-founder and CTO of MaintainX, a startup that has raised over a hundred million dollars in venture capital to bring AI to industrial operations in an intuitive way. We haven't had many episodes on industrial applications of AI despite industry generating oodles and oodles of data. Making this a great episode to hear about the enormous amount of opportunity applying data science to this space. Alright, enjoy. Here we go. Hugo, welcome to the Super Data Science Podcast. It's a delight to have you here. Where in the world are you calling in from today?
Hugo D.:	00:45	Hey, thanks Joh. Thanks so much for having me. I'm calling from Miami in the south of the us.
Jon Krohn:	00:51	Excellent. And like me, you're actually a Canadian originally, but you're from the francophone part of Canada from Quebec and you studied computer science, computer engineering at the university Sherbrooke in Quebec, right?
Hugo D.:	01:07	Yes, that's correct. So you computer engineering and then made my way to the US afterward.
Jon Krohn:	01:11	Yeah, a bit of a stint in the Bay area before founding your own company.
Hugo D.:	01:15	Company. That's correct. Was in the bay for about six years, founded maintenance from there. Access to network in the Bay Area is definitely something that is not present anywhere else and that's why I wanted to get to the Bay Area in the first place and then many during the pandemic made my way east. And then a lot of our company is based on the east coast, so actually it makes it very convenient. I can make my way to our Montreal

office frequently and it aligns all the time zones, so it makes it a lot easier day to day.

- Jon Krohn: 01:42 Fantastic. Yeah, that Bay Area network in tech in general and then in AI specifically, it is unrivaled. There's nothing like it. I'm certainly, every time I visit there I'm like man, I really should have considered moving here or should move here at some point. But yeah, I don't know. I'm an east coaster somehow end up in New York, but yeah, Miami as well where you've been drawn to now obviously the weather particularly in the winter. Fantastic. And so another place that calls my name sometimes, but let's talk about your company. So you founded Maintain X in 2018 after that Bay Area stint and you've been CTO there ever since you are. Your website describes maintain X as the undisputed leader in CMMS and obviously that marketing message is targeting a different person than me because when I go and I see that on your website, I'm like what's A-C-M-M-S? So Hugo, fill us in.
- Hugo D.: 02:45 Yeah, for sure. CMMS stands for Computerized Maintenance Management system at the end of the day. What does that mean? Very simple. It's work execution software for industrial and manufacturing. You can think about, it's a digitized version of a clipboard. Everybody in the plant going around with clipboards that have procedures, inspection, all kind of different things that they need to fill. We provide a digital version to get reports and be able to connect to machines as well so that we can gather more data and bring more insights.
- Jon Krohn: 03:13 So when you say that it's kind of like an iPad or something like that, like a tablet that people walk around with or is that just an analogy?
- Hugo D.: 03:21 Yeah, that's more of an analogy. So for us we have a web app that people can use on their computer and we have a mobile app either on a phone or a tablet depending on

what they decide to use. But really that's where the analogy comes in where on the app we have ways to build procedures, we have ways for technicians to go around and fill them and we want to make it feel like what they're used to because at the end of the day, change management is probably one of the hardest problems. So how do we reduce that? And it's making it feel like what they're used to but just in a digitized format and make it easy for them as they're going around.

- Jon Krohn: 03:56 Gotcha. So in any kind of industrial setting, like say I'm a manufacturer and historically I've had a bunch of employees keeping tabs on how things are going on my manufacturing floor. Historically they would've been walking around with a clipboard and just writing down notes manually in their probably horrible handwriting, they would've been just putting check marks in boxes because I suspect there's a lot of process of making sure that everything's going smoothly. So you're kind like, here's my checklist and you just go through and you walk around the floor and check the checklist. It's probably something like that historically that now you've replaced with an app.
- Hugo D.: 04:35 Yeah, a hundred percent. Most customers that we see oftentimes came from paper, so they were printing work orders. Either they had them in their own custom work orders in Word, they might have had some solution that would allow to print, but we in the field and then we've seen customer tell us to handle high priority work order. For example, they have a red stack of paper right next to the printer, they change the paper in the printer and if you see someone going with a red sheet around the plant, that means something is really high priority. That's the industry we're coming in and trying to really help and make sure that we have ways to capture that data, so much valuable data that they're capturing every day. And unfortunately it was in file cabinet all along

- Jon Krohn: 05:15 And so you guys have had a lot of success to be able to claim that you're the undisputed leader in this space and CMMS. That's tremendous. And so it sounds to me like a big part of your success was creating an application that looks and feels to these workers, to these people on an industrial plant floor to make it feel like a consumer grade application to make it feel like they're on LinkedIn or Amazon or any of these other consumer grade applications that people are used to. Is that right?
- Hugo D.: 05:52 Yeah, that's correct. Growing up, we always heard enterprise grade, enterprise grade everywhere and when my co-founder and I came into the space and realized what enterprise grade software looked like, we were like, okay, there needs to be a massive change here because all the big social media company, they got to billions of users without any training. Why does it take eight weeks to train a new worker on their maintenance application In the world that we are right now where personnel is retiring at an alarmingly fast rate where there is people that are new to the field, machines aren't getting more complex, there's no time for us to take eight weeks to form or train a new worker. You need something that people can just pick up and they just understand it. And so for us it was mimicking a lot of the consumer patterns, whether it's in the way that app is navigated or the inputs or the way the screen feels, you really want to bring something that feels familiar so that they can pick it up right away.
- Jon Krohn: 06:50 Makes perfect sense. I hadn't thought about it from that perspective of intuition where with most consumer grade applications, the expectation is that you can just pick it up and use it and it's just you're going to be able to figure out all the bells and whistles kind of in an organic way instead of needing, like you described there, eight weeks of training, which I guess was the standard in CMMS before maintain X came along. That is definitely an

interesting point. And then so how do you work in things like data and ai? How do you integrate those kinds of capabilities into your maintain X platform and maybe to kind of color to give us some color, maybe you could walk through a common user story of a user of maintain X that involves some kind of AI element.

Hugo D.:

07:41

Yeah, for sure. I think to pre-phase this, I think just to talk a little bit more about data, when we talk about industrial data, there's different sources of data that come into play. The first and foremost sources of data that's the most common will be just human generated data, like technician walking around the plant, doing repairs, putting data in the system, creating reports. We're still in the world where a lot of the machine have maybe sensors but they're not connected yet. And so that's why we invested a lot in making it as easy as possible for people to be able to come in input data in the platform and making sure that we capture all this information that was in people's head before. We have technicians that have been there for 20 years, 30 years, and they have all this knowledge that has been accumulated and when new technicians come in, they cannot access that information.

08:31

So we're really trying to find ways whether it through voice typing to make it as easy as possible and then you can get into the next level of maturity, which is all those machines have a ton of sensor on them and there's a lot of interesting information from temperature to vibration analysis to current data. There's so much we can do with this. And so it's then capturing all this information and making it accessible because from vibration data, you can actually know if a machine is about to break, you can know what type of breakage x-ray is about to happen and such, just knowing that it's about to break, you can see if it's a misalignment, you can see if it's problem with the oral. You can find all kind of different issues from this

information. And so when we're thinking about data, for us it's like the capture part is probably the most important piece.

- 09:20 I feel like oftentimes people talk about data and they assume a state where all this data is already pretty accessible and on the platform, but when we think about manufacturing, one of the big gaps that we've seen is that we have to make sure that the machine are instrumented in the first place that users are able to bring this data in. After that then it's like what can we do with it? Right? So first part I mentioned predictive maintenance is a super interesting use case with machine information. We can detect fall, classify them and get into prescriptive maintenance, which is kind of like the holy grail of maintenance, which is how do you take the fault information, mix that with work order history with the manuals of the machine, with maybe OEM, the manufacturer's information as well, and be able to tell the user what step they should be taking or recommend actions of when and what they should be doing to make sure that this machine doesn't go down.
- 10:17 So that's one use case. Second use case that we see in the field as well. At the end of the day right now there's a mechanism study that says that about 60% of the technician's time is spent on researching troubleshooting and writing reports. So for us when we're thinking about AI is how do we reduce this? The first step is researching. Right now most of these machines manual are still on paper. So when something goes wrong with the machine and you see an error code that's like FD three, four on the machine and you're like, Hmm, what does that mean? Unless you've been in the field for 20 years and you've seen this code a hundred times, you have no clue You have to go back to the library, find a manual, go to page 175, find that error code and try to solve it.



- Jon Krohn: 11:01 When you say library, you mean a physical library of books?
- Hugo D.: 11:06 Yes. Internally in the company they'll have a book room with all their manuals just sitting there. That's actually one of the big problem we've seen with customers is how do we get those manual digitized and we've actually had customers so excited about some of the new A capabilities because one of the things that we offer is allowing people to bring those manuals digitally on the assets and be able to search through them with generative AI building kind of a copilot assistant for all this data. They've actually hired temporary staff over the summer to digitize those manual by hand and get them in the system. But yeah, this is where we stand in this industry. It's like a lot of this data is still on paper and so we really have to focus on how do we bring this data in an efficient fashion because I think there's so much we can do with it.
- Jon Krohn: 11:52 Wow. Alright. I think I might've interrupted your second use case there, but maybe you kind of got to it.
- Hugo D.: 11:57 Yeah, it was very much aligned with this, which is like, okay, everybody's walking with a supercomputer in their pocket all day, so how do we make this data accessible to that supercomputer? I think generative AI brings a lot of new capabilities on this front about summarizing, searching information that would've been really hard to find before. And so people can now be at the machine directly search to air code, get some information or ask directly, maybe they don't know what the name of that piece is, how often does it happen? You look at a car and you're like, oh, this piece is wrong. No clue how to even search for this. And so I think generative AI brings a lot of new capabilities to align language and making sure that no matter how you call it, they're able to find it and bring that information to you.



- Jon Krohn: 12:39 Nice. Well, fantastic work, Hugo. It's amazing the technology that you've built over these years. Tell us a bit about the tech stack that you use for building technology like this.
- Hugo D.: 12:53 Yeah, for sure. Our tech stack is fairly narrow because we want to make sure that our team can work on it in an efficient fashion, but we always invest where we need. So where main application tech stack would be mostly types, strip no gs, react native types strip, so a very aligned stack on the product side so that we can move quickly. And then when we think about data, we're paring a lot of our data insights through Databricks, so we're moving all our data into Databricks warehouse, Python heavily on the AI front. Rag pipelines are still one of the best tool that we have out there to be able to get insights and actually bring information to users. I think this is sometimes people will say, oh, rag is outdated, but to me it's still one of the best tool out there when you're, you're looking to bring information to people and you live in a world where there's a lot of data models have increased context, but it's not an easy solution. You can't just pass the data along the network all the time.
- Jon Krohn: 13:54 Yeah. When you're talking about situations like you were there with, once you digitize that library and you have all of those manuals converted into a digital format, rag is exactly what I had in my mind. Retrieval augmented generation because that seems like the perfect kind of pipeline for pulling out relevant parts of say a large number of books that you've digitized and then using some kind of chat interface, some kind of LLM to pull out the relevant information from those particular pages, from the particular books that you need and be able to answer somebody's question in natural language. I think it's a no-brainer use case for rag.

- Hugo D.: 14:36 Yeah, exactly. And that's where we see kind of the first use case, the first foray, because I think it also, it's about helping the user educate themselves about how to leverage ai. It's new for everyone and we want to make sure that we make it transparent to them and easy to understand what are the implications of what they get. And then the part we're investing now is around background agents has mentioned a lot of the time is spent on researching but also writing reports. So I feel like agents are a great use case that can work in the background once you finish a work order that can work for you and bring that report to the finish line and then let you approve it so that it's sent to your manager automatically. A lot of the work also is like parts inventory management going on five website, picking the best part at the best price is not something that you need a human to do, and I'd much rather the part manager thinking about is part strategy and think about what's the future and have agent that can work for them and really make sure that all this information is done in the background captured and then brought back to them.
- 15:38 And I think this is a case where agents can really come into play and it gets really interesting. We're seeing a world where agents should be kind of colleagues to the technician where all these kind of jobs are where we automate the non-human work out of the loop, right? Human should be doing human work and so all of this repetitive work can we have an agent that's able to do it and human is able to orchestrate this is capabilities that's coming oftentimes I feel like in the tech world and in a lot of tech companies and are available but haven't made their way in the manufacturing world yet. And I think this is where the big opportunity is around ai.
- Jon Krohn: 16:15 Excellent. Then, so when you're talking about things like the predictive analytics that you were doing, so you're talking about how critical it is to be collecting data on

vibrations, say on some piece of equipment so that you can predict when some piece of equipment is about to fail instead of after it's failed, which I can imagine for your clients is a huge money and time saver because a piece of equipment actually going down as opposed to being about to go down and fixed it can make a huge difference in terms of downtime in say a manufacturing setting. So I can see how valuable that is when you're building those data pipelines or the analytics maybe building an AI model to do predictive analytics. This might be proprietary, so maybe there's nothing that you can disclose here, but if you're able to, what kinds of frameworks are you using there? Is it Databricks alone or are there kind of things like psychic learn involved, other kinds of regression modeling technologies or something like that?

Hugo D.:

17:19

Yeah, it's a mix to be honest, because one of the main challenges, you don't have the same amount of data across every customer. Some customer, you work with them and they have three years of historical data. You can train some pretty advanced models there. You can go into deep learning techniques, which allows for more cross machine training instead of being one model, one machine, but not everybody is there, so you have to think about how do I get started quickly? And oftentimes some of those simpler techniques will work really well when you have limited capability of limited amounts of data and so that you can at least get going and provide some value, right? At the end of the day, what's most important is time to value. And so we use different techniques depending on the amount of data, and that's something that we're refining as well so that we can have a maturity curve of having maybe some more generic models, simpler that works with you at the beginning, and as you progress, we can train some more advanced models with more advanced techniques that really brings you the deeper insights on this front.

- Jon Krohn: 18:25 Nice. Thank you for that answer for giving us some insight behind the curtain. And then in terms of, this is a question that I know you can kind of go into a lot of detail on, and I'm not asking for any kind of secret sauce here, is when you are building a tool like this that is supposed to look and feel intuitive when you are building some kind of predictive or AI capability into that tool, what kinds of considerations do you need to have to ensure that a non-technical user, a lot of our listeners are technical hands-on practitioners, but if you have somebody that's used to having a clipboard on a manufacturing floor and now they have all this AI functionality within an application, what kinds of considerations do you have to make that AI intuitive, intuitive for the users and just feel like a good experience?
- Hugo D.: 19:19 Yeah, no, that's a great question and I think that's something we've been spending a lot of time. So couple items related to this. One, it's making sure that you are clear when there's something that is AI generated you working with, you want to make it clear when something comes from AI so that you have clear sources as to where that information comes from so that it can be DoubleCheck In a lot of cases if you're writing content or emails with ai, if some of the content is hallucinates a little bit, you're not in a dangerous situation, but people are working with machines that have high voltage, high frequency, a lot of things can go wrong. You want to make sure that when you are recommending steps or action that somebody wants to take, that it really is grounded on manual information or ative information. And so we have a lot of checks and balances in place to make sure that we reduce hallucination.
- 20:14 We have multiple models that will check the answers as also afterward and validate what was generated. But then the final step is really making it transparent to the user, really being clear with them that this is something that

comes from AI and here's the sources so that they can double check and validate. Because at the end of the day, safety is really, really important to us and it's important to our customers and we want to make sure that we keep that in mind. So that's probably the first principle.

Second, it's making sure that AI will be where you expect it to be. What I found from the best tools that I've used on the software side is that maybe it's the settings or the button or the action that I want to take next is always kind of in my workflow and in line, not something that I have to go check three menus down the line. And so it's really thinking about how do we make those call to action with AI as within the workflow as possible. And so there's no secret sauce for that. It's really sitting down with a user talking about their workflow, how do they do things, where do they expect things to be, and really making sure that it feels natural as they're progressing.

Jon Krohn: 21:18 That is essential. This idea of having the AI tool be in the workflow feel integrated, that's the key to success. I agree a hundred percent. Nice. Now, so one last technical question that I have for you, Hugo, is I understand from speaking with you before recording the episode that there's a bit of a paradox in manufacturing data where manufacturing produces more data than any other sector by a huge amount, but when people are in a manufacturing plant, it often feels like a low data environment. What's up with that?

Hugo D.: 21:59 Yeah, yeah. There's some studies that says that manufacturing alone generates about 2000 petabytes of data per year. That's a lot of this data when you think about it comes from the fact that it's not human generated, but it comes from it's machine generated similar to in tech, like there's a lot more data coming from server logs and things like that than there can be from humans doing actions to the software. So with machines that are working 24 7, there's a lot of information you can

get from all kinds of different sensors that are on that machine. The main problem is oftentimes this data is not going anywhere. It's going into a void. Either the machine is instrumented but the data doesn't get anywhere. Sometimes they'll have a historian and so it's running onsite. They have something that aggregates the data, but then nothing is done out of it.

22:45

And so I think that's one of the biggest part for us is how do we make sure that we can bring this data to the cloud and after that make it in an actionable fashion. I've seen a lot of software that go into predicted maintenance and their target persona will be the technician or the maintenance manager, but then they will showcase really advanced vibration analysis charts and not everybody is formed or trained to be able to read those charts. So one of our concern all the time when ingesting this kind of quantity of data and showing it back to the user is how do we make it as simple as possible to get to the insights? It shouldn't require a degree to be able to read the data that's in system. Sure, we can make that available for the few people that have the certification and are trained to read this, but we need to make data accessible. And I think that's one of the biggest gap in the space is that they have all this data coming in, but they don't have that analysis part that really makes it actionable. Insights for technicians and the maintenance team,

Jon Krohn:

23:44

Plenty rich opportunity for more data analysis, more AI for maintain X and other folks working in the industrial space. What a great spot for you guys to be in to already be established to have this 7-year-old company in this space. And as we collect more and more data from more and more sensors, how well positioned you and maintain X are to capitalize on that and continue to build intuitive tools that allow your customers and their workers, their users to be able to glean insights and make an impact in ways that previously were never possible.



- Hugo D.: 24:24 John, I totally agree with that, and I think this is a super exciting time. We're seeing a lot of nearshoring of manufacturing as well. It brings so many new opportunities, so many new plants coming online as well. Automatization also is increasing. We're seeing more robots going around. That just means more maintenance as well. Those machines don't maintain themselves. And so I think it's a super exciting time to build and this industry is just continuing to grow right now, and I see so much potential for it in the future.
- Jon Krohn: 24:53 Fantastic. I'm going to blindside you a bit here, Hugo, but because I usually warn my guests in advance before I ask them this question, at the end of every episode, I ask my guests for a book recommendation and I forgot to warn you about that, but you're smiling, so it looks like you're not too intimidated by that.
- Hugo D.: 25:12 Yeah, no, I think there's a lot of good book recommendation. It's funny that you mentioned that because maintain scoring quickly, we do a lot of interviews and that's one of the questions I really like to ask every manager that we hire as well. And so I think turn to Ship around is probably a really good book around fast grow companies. This is something that we've actually had on the docket for book club recently as well and was really liked internally. So I'd say that's definitely one I'd recommend and then maybe crucial conversation because at the end of the day, no matter the amount of data automatization that we have, a lot of the problems that are in companies are human problem. And I think communication is still by far one of the biggest challenge and so crucial conversation I think does a really good job addressing this.
- Jon Krohn: 25:59 For sure. Thank you for those great wrecks. Turn the ship around in crucial conversation. So you guys have a book club at Maintain X?



Hugo D.:	26:05	We do. We do. And it's open to all engineering, and so we have books every two months because at the end of the day we don't want to rush too much, but where people will read a similar book, we buy copies for people internally that are interested and then we have a discussion around it.
Jon Krohn:	26:20	I love that. That's a great idea. Something I should institute in my own businesses. Fantastic. Hugo, it's been so great having you on the show today for people who want to follow you after this episode, where's the best place to do that?
Hugo D.:	26:32	Probably on LinkedIn at the moment. That's probably where it poses the most content, to be honest. In the past seven years, I've been very heads down billing products, so I think LinkedIn's probably the best spot at the moment.
Jon Krohn:	26:45	Makes a lot of sense. That is the number one answer that I get from my guests by far. That is the platform that we all seem to have ended up on now. Hugo, thanks so much for taking the time with us today. Such an interesting episode. It's nice to get some insights into how data and AI are being leveraged in the industrial world, something we don't do enough of on this show. So greatly appreciate it
Hugo D.:	27:09	A hundred percent. Joh, it's been a pleasure and I hope you have a really great day as well.
Jon Krohn:	27:15	So cool to hear about how MaintainX is bringing modern data and AI to manufacturing. In today's episode, Hugo Dozois-Caouette covered the three key sources of industrial data, human generated reports, machine sensor data, and digitized equipment manuals that enable predictive maintenance. He talked about how rag pipelines and AI copilots help technicians instantly



search error codes and troubleshoot issues instead of hunting through physical manuals and company libraries. He talked about the development of background AI agents that automatically generate reports, manage parts, inventory, and handle repetitive tasks while humans get to then focus on strategic work. And he talked about why transparency and safety are paramount when implementing AI in high voltage industrial environments, requiring clear source attribution and multiple model validation checks. Alright, I hope you enjoyed today's episode. To be sure not to miss any of our exciting upcoming episodes, subscribe to this podcast if you haven't already. But most importantly, I hope you'll just keep on listening. 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.