



SDS PODCAST EPISODE 59 WITH DAN SHIEBLER



Kirill: This is episode number 59 with Data Scientist at True Motion Dan Shiebler.

(background music plays)

Welcome to the SuperDataScience podcast. My name is Kirill Eremenko, data science coach and lifestyle entrepreneur. And each week we bring you inspiring people and ideas to help you build your successful career in data science. Thanks for being here today and now let's make the complex simple.

(background music plays)

Hey guys, welcome back to the SuperDataScience podcast and today I've got a very, very interesting episode for you. Today on the show I had Dan Shiebler, who is a data scientist at True Motion. So first things first. True Motion is a company that develops these apps which you can install on your phone and you take them in with you into your car, take your phone into the car, and it will automatically detect that you're driving, if you're the person driving or if you are in a passenger seat. It will also detect how well you're driving, if you're being distracted while driving, if you're driving too aggressively, if you're following certain safe driving techniques or not. And it will rate your progress on that.

So how cool is that? All that happens automatically, you don't even have to interact with the app, it does all of that in the background. So if you want to check out True Motion, then go to gotruemotion.com/app, you can download it there completely for free and test it out for yourself. And today we will understand how all this works thanks to Dan, who is



one of the brains behind this whole operation, who does the data science of this app, and he will be on the show today.

So Dan came on and shared a lot of insights about how sensor data from your phone can be collected and analyzed and how data science can be applied to that to generate valuable insights. So you'll definitely find out more about that. We'll talk about Dan's background as well, and how he got into data science and how he ended up working in this amazing and exciting space. And probably one of the coolest parts about this podcast is that we're discussing very applied things. So you will actually see how data science is used in action, how data science enables this completely new industry, this completely new way of thinking about driving, and how data science changes lives. It's a very, very important thing to be able to actually see the value that data science can bring, and that's exactly what Dan shared with us today. So without further ado, I bring to you Dan Shiebler of True Motion.

(background music plays)

Welcome everybody to the SuperDataScience podcast. Today I've got a super exciting guest, Dan Shiebler. We met with Dan at the ODSC conference, and he was giving a very exciting talk on applied machine learning, so I'm super excited to have Dan on board. Welcome to the show, Dan. How are you going today?

Dan: I'm doing great. Really happy to be here.

Kirill: And where are you calling in from?

Dan: Calling in from Boston, Massachusetts.



- Kirill: Aw, fantastic. That's exactly where the conference was and where I attended your talk. So how's the weather in Boston right now?
- Dan: Terrible, to be honest.
- Kirill: It's like always like that, isn't it? When I was there, it was raining for 5 days.
- Dan: Sometimes it clears up and you appreciate it so much more.
- Kirill: Have you been to London, by any chance?
- Dan: Once or twice, I have.
- Kirill: How does the weather compare in Boston to London?
- Dan: Remarkably similar! They called it New England for a reason!
- Kirill: That's the reason! I was like, why did they call that part of the US new England? It was so confusing! And how does it work, tell me before we jump in the podcast. How does it work with football teams? You only have one? You only have a New England team? Or you have several?
- Dan: No, there are several. There is the New England Patriots, is the main football team in New England. I'm actually not sure if there is any other New England football team. Patriots are pretty dominant.
- Kirill: Good, yeah, I've definitely heard of the Patriots. Ok, ok, we're getting a bit sidetracked. Thank you so much for coming on the show. Your talk was very inspiring, one of the most practical talks, and I just couldn't wait to get you on here to share with the world about how data science and machine learning can really be used in the real world. So how about



we start with where you work. What's the company called where you work and what do you do there?

Dan: I work at True Motion. We develop smartphone apps that can track how good of a driver you are and we do this in order to make driving safer. We work with some of the largest insurance companies in the world in order to give people discounts on their car insurance and incentivise safe driving. My job is developing machine learning algorithms that operate on smartphone sensor data in order to track how good of a driver you are, identify dangerous driving behaviours, and suggest ways that you can improve your driving.

Kirill: That's so cool. I've heard of this, it's not a big thing in Australia, where I am, or maybe it is a thing, but I don't know if people use it or not at all. But is it compulsory in the US, or do people sign up to it voluntarily?

Dan: It's not. It's actually been growing for the past 10 or 15 years. There have been programmes in place where you would install custom hardware in your car and it's sort of a pay-as-you-go insurance where if you drive more and you drive less safely, then your car insurance will go up, and if you drive more safely or drive less, your car insurance will go down. So it helps people who feel that they shouldn't be getting charged as much as other people and get discounts. And we're sort of bringing it to the masses by moving the technology off of customized hardware you have to install in your vehicle to just an app that you can download on your phone. It's live in about half of the US states at the moment, and it's growing.



- Kirill: That's really cool, and I just wanted to reiterate that for people listening, I've heard of the custom hardware you install, but this is so cool, this is just like on your phone. You just put it in your phone and of course, there's like lots of complications with that, and we'll get into those in a second. So when you download this app, does a person have to register with your company, or does their insurance company already have to support your app?
- Dan: So we have a number of different programmes that we offer. There's some programmes we go directly through insurance companies, where when somebody signs up on an insurance company, they would download the insurance company's app, which would have our software and our algorithms running on it. We also offer some of our own commercial apps. We have TrueMotion Family and Mojo, which are different applications that have the same algorithms and same idea. One is for parents and teens to be able to learn how to become safer drivers together. That's TrueMotion Family. And Mojo is an app for people to become rewarded for driving without being distracted. You accumulate points based off of how many distraction-free miles of driving you have. And when you get enough points, you are eligible to win Amazon gift cards.
- Kirill: Oh, that's so cool. That's your company's contribution back to the community, right?
- Dan: Exactly.
- Kirill: That's awesome. This is the next question I had: What is defined as distraction-free driving and what is safe driving in general, according to your definitions?

Dan: Again, this is sort of an application-specific question. In terms of distraction, we are really focused on distraction as being interactions with your cell phone. And there's a variety of different types of interactions people have with their cell phones while they're driving, some of which are more or less dangerous. There have been a number of longitudinal studies that have been performed on thousands of drivers with cameras in people's cars that have really broken down what types of distraction actually increase people's risks of getting into accidents by just tracking what types of distractions are people doing and did they get into accidents.

Really, the bottom line is the type of tap-on-screen and swiping and sending text messages and active interaction with your cell phone is very dangerous. Whereas using your GPS app or listening to music with your cell phone, that's relatively safe. From the perspective of the car itself and the other safe and dangerous driving and things that people do, one really strong indicator of whether somebody is driving safely is their braking habits. When somebody brakes really hard, that's usually indicative that they were driving at a speed that is above that of those around them or that they were driving aggressively and sort of cutting in and out of lanes and such, trying to get ahead and then they had to slam on the brakes for one reason or another.

Kirill: Yeah, I totally agree. I'm sure all of us have been in situations where there's like a driver in front of us braking very hard on a highway or something and then you have to brake very hard as well because of that person or somebody is cutting in in front of you. I definitely see where that is coming from. So, your talk was titled something along the

lines “The Number of Sensors in Your Phone” or “The Variety of Sensors.” Can you correct me on the title and let’s talk a bit about that? How many sensors exist in a modern phone?

Dan: The title of the talk was “The Power and Pains of Sensor Data.” Like you’re alluding to, there are a very wide number of sensors that are in your phone and in other devices that you surround yourself with every day. I don’t know if I have a number right off the top of my head, maybe around 10-15 different useful sensors that would provide rich time series data that you could use, just from your cell phone.

Kirill: Gotcha. Give us a couple of examples. I’ll start with probably the ones that everybody knows. Like, a GPS, a temperature gauge and maybe like a barometer. Those are the first ones that spring to mind. What else do you have in mind for a phone?

Dan: Probably the three of the ones that we think about the most, in addition to the ones you’ve mentioned, would be the accelerometer, the gyroscope and the magnetometer, which all sort of fall into the general set of the IMU, which can track the position of the phone and the sorts of forces that are currently acting on the phone. And that’s really useful for determining where the phone is, what the phone is doing, recognizing people’s activity, and from our perspective, recognizing vehicle activity.

Kirill: Let’s go through a couple of those. You said — temperature I understand, GPS — what is GPS in your phone? It’s like a little microchip, right?

Dan: Yeah, it communicates with satellites. It uses the Doppler Effect to get an estimate of your current position and an

estimate of your speed. Actually, the speed estimate from the GPS is more reliable than the position estimate because the signal that's being measured directly is the speed and the position sort of being computed from that.

Kirill: That's cool. Barometer tells you how high you are, how uphill or downhill you are, is that right?

Dan: Barometer is the pressure sensor. So it can tell the air pressure around you. You can use it as a proxy for altitude, because the higher you go up, the lower the air pressure is around you. And you can use the barometer for other pretty cool things. You can see on the barometer signal when someone opens up a window in their car or something like that.

Kirill: Yeah, the pressure in the car changes, right?

Dan: Yeah.

Kirill: How does the accelerometer work? I'm really curious about that one.

Dan: The accelerometer measures force and acceleration. Different accelerometers work in different ways. I believe the one in the cell phone works with a capacitor. It's got three axes, so really it's three accelerometers that are fixed in the X, Y, and Z axes of the phone, and when force is exerted on the phone, it creates a spike in capacitance that generates voltage signals that the phone then can interpret as accelerations along the different axes.

Kirill: Okay, that's really cool. Okay, the next one is gyroscope. What does that one tell us?

- Dan: The gyroscope picks out rotational acceleration. Probably the best way to think about the gyroscope and the accelerometer is that the accelerometer pulls out linear accelerations and the gyroscope pulls out rotational accelerations. That would be like turning along the axes of the phone. If you think of spinning your phone in your hand or even spinning it on the table, those are the sorts of things that the gyroscope would pick up.
- Kirill: Okay, gotcha. And that would be helpful for you to understand if somebody has picked up their phone or not.
- Dan: Yes. It's also helpful to understand what the orientation of the phone is and how the orientation is changing in space.
- Kirill: Gotcha. We'll get to that in a second. Let's talk about the magnetometer. What do you use that for?
- Dan: The magnetometer is—sort of canonical use of it is for a compass. It picks up the magnetic fields and lets you orient the phone with respect to the Earth. In turn, when you integrate the magnetometer with the accelerometer and the gyroscope, it gives you a third party reference for the phone's position. With just the accelerometer and the gyroscope, if you're trying to track how the phone changes position over time, you experience this sort of drift that occurs as errors accumulate and you lose track of where the phone is exactly. But if you have a third external reference to the outside world, which actually could be the GPS or the magnetometer depending on your application, you can get a better sense of absolute location and minimize drift. Although in practice, being able to determine the phone's

position and see long-term position changes just using IMU is really difficult.

Kirill: Gotcha, yeah. Thanks for that overview. People who are listening to this podcast, you might be wondering why I ask all these questions. The reason I ask is, as with any data science challenge, it's important to understand what input data you have. First of all, of course, it's very interesting that you have these six and many more very powerful sensors in your phone which you probably didn't even know about, but at the same time it's very important to know what input data you have to deal with when you're solving a data science challenge, especially something real world like this.

So, Dan, tell us more. Now that we know what data you have, what do you do with this data to come up with those insights that we're talking about to tell if a person is driving safe or they're driving risky?

Dan: There's a wide range of different algorithms that we would want to do and different sorts of pre-processing that's useful to operate on for each of these algorithms. A good example would be a type of pre-processing we want to use with all of our different algorithms, which would be pulling out the gravitational acceleration of the phone. What this means is the accelerometer lets you see what the linear accelerations on the phone are, but understanding which of those accelerations are from gravity and which of those accelerations are from other sources lets you find exactly what is the orientation of the phone, which could tell you if the phone is in a pocket, if the phone is in somebody's hand, if the phone is in a mount, if the phone is flat on a surface.

That gives us a lot of insight into what the phone is doing at a particular point in time. And in order to do that, we would use a sensor fusion algorithm. A typical one is the Kalman filter to fuse the inputs from the gyroscope and the accelerometer in order to pull out those gravitational accelerations to get an estimate of the phone's position based on that.

Kirill: Okay, gotcha. Basically, if my phone is in my pocket in the car, if it's in the car door, or if it's on the seat next to me, or if it's in my hand, you can tell that through this sensor fusion approach that you're taking to these sensors.

Dan: Yeah, exactly. And once we have that signal, it becomes way easier to tell all the other things we want to tell like pull out when a car is doing a turn or pull out when someone's got the phone in their hand because we've got this more reliable, derived signal from those sensors. This is a pattern that pops up a lot in all of this sensor processing where you have the raw signal. It's usually difficult to build a machine learning algorithm to operate directly on that, but there's this 100 years of signal processing algorithms and mathematics that have been developed that it's really useful to form these pre-processed signals from your raw signals before using them in your machine learning algorithms.

Kirill: I understand. I just wanted to also highlight how important this is. For me, this was one of the key points in your talk, really revelational that you cannot just use the sensor data as it is. You have to understand a very important thing, and that is how the phone is oriented in the car to understand which sensor corresponds to braking, which sensor really

corresponds to centrifugal force when you're going around a corner and those things. So you have to kind of prepare the playing field in order to understand these things.

That was really cool that you guys have come up with a way to do that effectively and then actually use the data in the correct meaning that it is conveying. The next question is then, once you know all these things, and you know how the phone's oriented and you can derive these metrics from it, what kind of machine learning algorithms do you apply to that data?

Dan: One pretty cool thing—I think before talking about machine learning algorithm, maybe I'll just flesh out which problem we're trying to solve.

Kirill: Sure.

Dan: One problem that pops up a lot for us is determining what sort of vehicle somebody is in at a point in time given the sensors that we're seeing. Because if we see that somebody is experiencing a whole bunch of bumps on their phone and it looks they're speeding up and slowing down and all of that, we don't want to automatically jump to the conclusion that this person is a bad driver just because they were in a bus and that bus was driving somewhere and the bus driver was terrible. And even more difficult would be if someone is on a bicycle. The sensors from a bicycle, if we thought it was a car, then we might think, "Wow, this guy is really a terrible driver. Look at how much the phone is moving all over the place..."

Kirill: "...and how slow he's going," yeah.

Dan: Yeah, exactly. It's useful for us to build algorithms that could identify mode of transit. And even harder would be driver identification, determining whether somebody is a driver or a passenger in the car.

Kirill: That's right. That was a big part of your talk. How do you do that? How do you decide if a person is a driver or a passenger in the car?

Dan: We have a whole ensemble of different algorithms to do that. One that uses the sensor data pre-processing we were talking about is the exit window method, which is an algorithm that we use for determining the moment someone exits the car, and given that, what side of their car they exited on. In the U.S., people drive on the left side of the car. So if you're exiting from the left door, then you're more likely to be the driver, and if you're exiting from the right door you're more likely to be the passenger.

It's not always the case because sometimes people will sit in the back left seat, but it's a pretty strong signal. And if we can get that estimate of gravitational acceleration on the phone and get an estimate of the orientation to phone, we can use the gyroscope to tell which side of the car the person exited from.

Kirill: Wow! That is so cool. That's just mind-blowing. Like, people get out of cars every day. They don't even know that there is data that is being collected about that. Not necessarily everybody has your app for now, but still, that can be determined from the data on your phone. That's so cool. All right. So now we know one of the problems. Are there any other problems that you're trying to determine? So you're

trying to determine what type of vehicle, whether it's a bus or a car or a bicycle, if it's driver or a passenger, then you're trying to understand also if it's a passenger or a driver if it's in a car. Are there any other complications that you have to solve before you get to the data processing part?

Dan: One pretty key problem is actually just detecting a trip, detecting when somebody is going fast enough or moving in a way that we would think, "Oh, this person has begun a trip," which would be like they've started driving or biking or taking a bus or something like that. One naïve way we could do that would be, "Let's just constantly ping their GPS, full power," and if we think that they're moving fast enough for a consistent amount of time, then we'll say, "Okay, this person is in a trip. Let's run all our mode of transit and driver identification algorithms on it and determine what's going on."

But this burns a lot of battery, so solving this problem is actually a little bit more difficult than just relying on the GPS to throttle the person's phone and tell us exactly what's going on. So we can use cell phone towers and Wi-Fi signals to triangulate somebody's position and determine whether or not they're moving. If the closest cell phone tower to somebody is changing at a certain pace such that the cell phone signal from one tower decreases and another one increases, and the same thing with Wi-Fi — Wi-Fi signals from certain Wi-Fi's start changing at a cyclical pattern. That's usually indicative of motion, especially at a particular speed. And when you see that signal, then you can decide, "Oh, let's turn up the GPS." So, processing and



understanding those signals from those sensors is a tough problem in and of itself.

Kirill: That is another really cool thing. It reminds me of — not “Mission: Impossible,” it was like — what’s that movie with Bruce Willis? They have like a series of those movies...

Dan: Which one? “Fast and Furious”?

Kirill: No, Bruce Willis.

Dan: That’s — “Die Hard,” maybe?

Kirill: Yeah, “Die Hard.” Like, “Die Hard 3” or “Die Hard 4” where they’re like in this big van and they’re travelling across the U.S. and they’re like stopping the whole government from that van, like stopping all the traffic lights and stuff. Nobody knows where they are and then the guy is like, “We can triangulate them. We can find out where they are through triangulation.” That’s like, “Oh, that’s what Dan does. Okay, got it.”

Dan: Exactly. Every day. Just like that.

Kirill: That is so cool.

Dan: We’ve got a set of machine guns and all that.

Kirill: (Laughs) That’s not a thing in Boston, in Massachusetts, right? That’s more like Texas, a set of machine guns and all?

Dan: Yeah, less popular here.

Kirill: (Laughs) Okay. All right, gotcha. So we’ve got three issues that we found out now. Whether a person is on a trip or not, what type of transit, and whether or not they’re a driver. All

right, now can we get to the algorithms or is there more challenges that you have to solve before you get there?

Dan: Well, there's always more challenges, but I think that's a pretty good overview.

Kirill: All right, cool. So, let's get to the algorithms. What type of machine learning approaches do you take to get the insights that you're after from this data?

Dan: If we're trying to determine how good of a driver somebody is, once we've sort of identified, "All right, this person is on a trip. Let's run our machine learning algorithms on them to pull out events," usually the way that we think about driving is in terms of discrete events. When somebody commits a hard braking event or commits a distracted driving event, then we have that [indecipherable] and then we'll aggregate those events and run a different algorithm on the aggregated events to get an estimate of risk. It's sort of a hierarchical pattern rather than an end-to-end model that runs directly on the sensors. That's the situation when we would be running a little bit more complex machine learning algorithms that operate directly on the sensor data.

For example, for pulling out distracted driving events, we would look at the signal and segment the signal into different short segments to try to estimate the moment when somebody begins a distracted driving event, when they pick up their phone or when they first unlock the screen. We have actually a few different models for different situations, whether the algorithm is running in the back end or running on the phone. Some are Random Forest and some are neural

networks, that would operate directly on the sensor data in order to pull out those events.

And once that model gets run on the trip, we would look at how the events that are predicted sort of look over time when we pass another model over them that reasons about the time series nature of the data. We have like a hidden Markov model that would reason about, “Well, if this person was using their phone at this point, maybe they’d be more likely to use this type of phone usage here or that type of phone usage there.” The models would want to produce different types of outputs in kind of a multi-class classification sense for whether the phone’s in the person’s hand and they’re swiping, or in the person’s hand and they’re tapping, or in the case of somebody on a phone call or something like that. I think that’s basically the algorithm overview for distracted driving case.

Kirill: Okay, gotcha. So you have two separate parts of your analysis. First of all, you apply some algorithms like Random Forest or a neural network directly on a sensor data to understand, to pull out those points in time where you have distracted driving events. And then, as I understood, you take those individual potentially distracted driving events and you put them into a time series and then you use a hidden Markov model on top of them to analyse what type of distracted driving behaviour that could have been.

Dan: Yeah, exactly. And we interject, at both points in the model, we interject a bit of other information that comes through. The operating systems of the phones give us some information on whether the screen was currently on or

whether the person was in the process of a phone call or placed a phone call. Those would help tune the behavioural modelling, hidden Markov model aspect of the algorithm based on the phone's inputs. And the key problem that appears there is the different types of phones, the Android, the iOS and the 10,000 different types of Android phones, give signals like that in very different situations and with different amounts of accuracy. The motion sensor signals are also different amounts of accuracy for those different types of phones. So we have to develop all of our algorithms in a way that they produce outputs that are really consistent between the different kinds of inputs that get provided by the different types of devices.

Kirill: Okay. You mentioned something around the legal side of things, that you have to make sure that your software works the same way on all the phones regardless of whether they're expensive, whether they're cheap, because you cannot discriminate against any people that might be using your software. Is that right?

Dan: Yeah. That's correct.

Kirill: Gotcha. That's an important and probably a difficult issue for you to solve since phones that are less expensive, they probably have less sensors. Is that what you find?

Dan: Yeah, absolutely. At some point we need to decide that certain phones we're just not going to support if we don't think that we're going to be able to grade people in a totally safe fashion. For example, some types of Android phones don't have gyroscopes. And if all of our algorithms are developed really relying on gyroscopes, and there's really

certain things you can only do if you have gyroscope information, you really can't include somebody in the program if they don't have a phone that has a gyroscope.

But beyond that, it really requires just an analysis of all of the different sources of data and doing sort of distribution analyses to make sure that the distribution of judgments for people, both from the raw sensor data, the pre-processed sensor data, and the responses that we produce, are consistent between the different types of devices. And we could use different kinds of distribution analysis methods like KL divergence and things like that to get those estimates.

Kirill: That sounds really cool. So that was the distracted driver approach and how you solved that. And before we proceed to the safe driver, just in a nutshell, very quickly, can you tell us what a hidden Markov model is?

Dan: A hidden Markov model is an algorithm for modelling time series based on the theory that at each time step there is some output that we've observed, but the system was in this hidden state that we didn't observe. At each time step, this hidden state progresses. That hidden state might be in this case that the phone is in the person's hand, the phone is down, the person is on a call. And when you're in a different state you have a different probability of emitting different visible signals, which could be the kinds of vibrations in the phone that would be indicative of these different situations. And you also have different probabilities of transitioning to the other states.

So when you suspect that your data at each point in time is reliant and will change based on what sort of state somebody is in and that their state transitions — state B is more likely to follow state A than follow state C or something like that. A hidden Markov model is a good choice. It makes some pretty strong assumptions, it makes the Markov assumptions, which we don't need to get into, but it's a pretty good model for those types of situations.

Kirill: Gotcha. Thank you very much for that overview. Okay, now let's proceed to the safe driver algorithm. How do you determine if somebody is a safe driver or not? What machine learning algorithms do you use for that?

Dan: One type of signal that's pretty reliable for picking out safe driving is the hard braking signal, which would be just pulling out the moments when someone slams on the brakes. We've got a few different models for that, actually, that we use in different situations. In fact, this is actually sort of unintuitive — the GPS signal, the speed signal for GPS is good enough that if we're simply trying to get an estimate of the moments when somebody brakes, if you just look at how the speed changes over time and use sort of advanced signal processing methods to get a good stable estimate of the derivative of the speed, you can pull out the braking periods without even using a machine learning model.

Just when you have that signal, it takes a lot more signal processing side in order to do that, though. The tough machine learning parts come in when you want to do the same thing without the GPS, because as we mentioned, GPS

really burns your battery. If you're trying to have the whole thing operate in a lightweight fashion where you don't have GPS throttling you and you want to just pull out the brakes on the accelerometer, the accelerometer's signal is a little bit too noisy to do it in a really simple rule-based fashion. So using machine learning algorithms that are pretty similar actually with segmenting the signal — you've got to do a whole bunch of pre-processing first – but then segmenting the signal and putting an algorithm on top of that works pretty well.

Kirill: Gotcha. And by this point, because you know the orientation of the phone, you can separate what is a braking signal and what is a centrifugal force signal meaning that the person is going around the corner. Is that right?

Dan: We can do that, but it's actually pretty difficult to get that signal separation just by using the gyroscope and the accelerometer. Maybe we could take a step back and sort of define this problem in a more complete way. The phone has three axes of acceleration, and the car has three axes of acceleration as well. For the phone it's the X, Y and Z axis and for the car it's the longitudinal axis, which is front to back; lateral axis, which is side to side; and vertical axis, which is up and down. And if we want to rotate the phone into the car's reference frame, we need to find a rotation matrix that would do that.

So just by looking at the accelerometer and the gyroscope and using the Kalman filter, we can resolve two of those degrees of freedom. We can rotate the phone so that it's aligned with gravity, but we still don't know the angle

between the phone screen and the front of the car. We have to use a different method to pull that out. That's the hard part.

To do that, we use principal component analysis, PCA. So as the car drives, the axes of acceleration that the car experiences will be longitudinal when the car speeds up or brakes, and lateral when the car takes a turn, because of the centripetal force. So since those are the axes of acceleration by which the car experiences most acceleration, those are the axes of maximal variance that the phone would experience acceleration on. So when we perform PCA on the motion sensor readings of the phone, after we've aligned it with gravity, that will then give us that angle that we need to rotate the phone so that it's aligned with the car and then we can pull out the brakes and turns and such.

Kirill: Okay. That's so cool. That's such a cool approach to that. Yeah, that's very interesting and now I can see how that works. And from that you can tell when the person is braking or accelerating too fast and there you go, you have your result that you are after. That's so cool. Thank you so much for the overview. It's really great to see how interesting this work is and what you're up to in the space of TrueMotion. I think this was really invaluable as a real practical application of data science and machine learning.

And what I would like to get into now is kind of move away a bit from the algorithms themselves and talk about your background, because a lot of people will be wondering how on earth does somebody get a job that interesting, a job that actually makes such a huge impact and is so much fun. Tell

us a bit about you. According to your LinkedIn, you went to Brown University, but there you studied something completely different. You studied neuroscience and also computer science. How did that go and where did that take you?

Dan: I started at Brown just studying neuroscience, actually, and pretty quickly got interested in the more computational side of neuroscience when I started doing some pretty cool neuroscience research on neurosurgery patients that would determine how people make decisions and drilling into sensor data, actually, understanding from microelectrode recordings in people's brains how different voltages fluctuate when people make different kinds of decisions and pulling out those sorts of signals. So really my first experience in data science was through a neuroscience lens and pretty similar in terms of the tools used and the techniques needed to what I'm doing right now.

I transitioned from that into a more computer science and computational focus. I really just loved how rich and exciting the field of computer science was. I loved the logic and mathematics part of it as well. And I think data science kind of came naturally from that. I originally came into computer science from sort of a data science perspective and that was the kind of focus that I had with mathematics and data science.

Kirill: Okay. That's really cool. So then after your degree, did you go straight to TrueMotion, or did you—it looks like you did some work at Brown University first.

- Dan: I did. I actually also worked for a short period of time at MathWorks, which makes MATLAB. I met one of the guys from TrueMotion and he told me about it and I just really couldn't resist. It sounded too exciting and too awesome.
- Kirill: That's so cool. So this work at TrueMotion, by the sounds of how you talk about it, it really excites you. Does it inspire you that you're actually making a difference to people's lives and that you're helping them achieve safer driving and better insurance policies?
- Dan: Yeah, it's cool to me that we can sort of build something and it will immediately be in the hands and changing the lives of so many different people. We've received a huge amount of feedback from different people. It's a very tight feedback loop on the algorithms I produced and it's really interesting difficult problems and it's nice to know that things that I do every day can help people become safer and help people save money.
- Kirill: It's always inspiring, I think, when you can see results of your work very quickly.
- Dan: Yeah, exactly.
- Kirill: Okay. In terms of the work that you do at TrueMotion, I know a lot of it is private information to the company, but is there a win that you recently had, something that you achieved and that you're really proud of, something that you can share with us?
- Dan: I don't know if I'd classify this as a win or something that we're really excited about in general, but I would consider it a win. We just finished building out our behaviour

modification platform, where we can run sort of really intense statistical analysis on people's behavioural change in real time and quickly change the experience that people are experiencing as such that we can measure the behaviour change in a statistically rigorous fashion on really short time scales.

So a lot of statistical analyses for people's behaviour and people's experience, it's really easy to screw them up and sort of stop tests early or put more people into a test after you've already checked its results and get p-values that you come out with and you really can't trust. We put a lot of time and thinking into building this model. We've spent a lot of time whiteboarding things out and running simulations and such to make it statistically rigorous. And we've got something that has got a pretty fast feedback loop that's changing people behaviours in a way that we can analyse it statistically rigorously and we're pretty excited about that.

Kirill: That's awesome. What do you mean by when you say you're "changing people's behaviours"? Like, is your app telling them what to do and how to behave?

Dan: Yes. The problem of behaviour change is pretty challenging with people's driving because people often perceive themselves as good drivers. We can give you a score afterwards and say, "You had three hard braking events and that means your score is terrible," or something like that. That will change people's behaviour, but not in a really good targeted fashion. So we're experimenting with lots of different regimes of push notifications, different presentations of the app, real-time alerts as well, like the

phone might ding you when you pick it up or hard brake or something like that, different kinds of patterns.

Really the kinds of user experience variations that most applications will have. Pretty much every company that makes apps or websites or something like that experiments with this at some point in order to change people's behaviour for one way or another, but for us it's a pretty high stakes game. If we change people's behaviour, we reduce car crashes. We're pretty excited about it.

Kirill: That's really cool. Love it. Love that example. Okay, the next one is—I don't know, maybe you've already covered this in what was discussed previously, but still I'm going to ask this anyway. What is your one most favourite thing about being a data scientist?

Dan: I think it's the way that data science changes really quickly. The fact that right now, this is sort of the very early days of data science and machine learning really in general, I think that there is this huge set of problems that have been sort of present and people have been aware of for a very long time. People have been trying to figure out ways to solve them and haven't really been able to for a while, but just recently people have been designing algorithms that use large amounts of data in order to solve problems that have been thought of as nearly impossible for almost a century. That's really exciting. We're kind of at these turning points in computational problem solving, and the way that these problems are being solved is with data science and with machine learning. So, being a part of that, being a part of this revolution in computational power, is really exciting.



- Kirill: That's so cool. I totally agree with that. And the other thing to that, the other aspect is that it's changing in so many different areas and so many different domains at the same time, like in medicine, in cars, in rockets, in mining, in social networks, everywhere. On that I wanted to ask you, do you ever feel that you're missing out? Do you ever feel that, yes, you're one of the top people driving the change in people's behaviour in driving and how to measure if they're driving safe or being distracted, which is a huge thing and has a huge impact, but at the same time, do you ever feel, "Oh, I wish I knew more about what's going on in the space of data science and medicine, or how data science is used in social networks." Do you ever get that feeling? Or are you set in what you're doing and you're just confident that you're happy and you're doing what you're passionate about?
- Dan: Well, I'd say, really, this is the information age. When I get curious about how data science is moving in another field, it's pretty easy for me to look up, read some research papers on what's going on elsewhere, and dive into it and work on it as much as I can. In fact, I've actually got a pretty good example. I've been interested in really the cutting edge of neural network models, convolutional neural network models, for image understanding and their relationship to neuroscience. There was a professor at Brown University who's done a lot of research on this, Thomas Serre, so I contacted him and I've been doing research with him in my spare time outside of TrueMotion, sort of working with him and some of his postdocs and grad students on some of the papers that they've been working on. It's been really cool being part of a cutting edge field of deep learning and data



science that's totally orthogonal to the sort of stuff I do at TrueMotion. I'd say there's opportunities to be involved in any of the kinds of problems that I'm really interested in.

Kirill: That's fantastic. I'm so glad you said that. A lot of people out there think that you really need to get a job and get into the industry. And that is correct, you do need to get a job eventually. But there's so many opportunities. Just like you gave this example, and it's such a great example. You're interested in the space of data science, convolutional neural networks, so you have spare time and you went and found a leading researcher in that field and now you're doing some work with him. That is such an inspiring example. I really hope people listening to this podcast will get inspired by that and get some ideas from that on how they can pursue something that they're passionate about, even if they're not necessarily doing work in that space right now.

Okay, that's really cool. And this all leads up to a question I'm really eager to ask you. From what you see about data science, from the work you do at TrueMotion, from the research that you do, from all these other interests that you have in data science and machine learning and neural networks and so on, where do you think the field of data science is going? And what do you think our listeners should prepare for to be ready for the future that's coming?

Dan: I think data science is going in a lot of different directions, but one thing that's really underlying all of it is this democratization of data science, which is just data science becoming easier for people to do, easier for people to learn, easier for people to do really powerful, really cool stuff with



cutting edge technologies without learning a humungous amount about all of the underlying math and underlying tech and just being able to sort of jump in and do stuff at an easier level.

And the reason why I think this, in addition to the huge number of research papers and automated data science and all of that that's being propagated, is if you think about programming computers, when computers first came out, and people first started programming computers, it was really hard. You had to use punch cards, then people were writing in Assembly. I mean, if people had to write in Assembly and never developed higher level languages on top of that, it would be really impossible to build the sorts of incredible things that people can build today.

It's really been possible because people have sort of put these extra layers on top of it that hide the nitty-gritty details and let people move building blocks around in an easier to understand fashion so they can bring their own expertise and own knowledge to build something really incredible. Because nobody can understand every single part of something really complicated.

We work by standing on the shoulders of other people's achievements, and I think that data science is going to work like that. I think, as data science becomes easier, it will be easier for people like doctors or physicists or such that have incredibly deep knowledge in their fields to use the tools of data science in a way that fits very nicely into the sort of work they've been doing already. I don't think this necessarily means that it's not useful for people to learn

about all the nitty-gritty details of data science. Just like any computer scientist can really benefit from understanding the fundamentals of a computer, I think that any data scientist can really benefit from understanding the mathematics underlying data science techniques. But I think that's something that we should be ready for. There's a lot of people coming into data science who don't have really firm mathematical backgrounds, and for data science becoming easier for people to use, easier for people to do.

Kirill:

Thank you so much. That's a very refreshing view on what is going on. Because we've had different opinions on the podcast. We've had people say that automation is going to completely drive the human aspect of data science away. We've had other opinions. Your opinion is very powerful and inspiring as well because that's what I think also, and that's how we structure our courses, in the sense that — just as an example, we try to avoid the complex mathematics and focus on the intuition and the practical side, because as you say, you don't really need to go into the nitty-gritty of what's going on in that algorithm to be able to understand on an intuitive level how it works and also be able to apply it in practice. That's an essential part, what matters.

We've spoken about cars quite a lot, but it's like driving a car. You kind of know what a car does when you accelerate and when you brake and things like that, and you know what approximately is going on inside the car and you know how to use it. You know how to apply that car to get you from A to B or to drive somebody else and things like that. You don't necessarily have to be a car mechanic and get into the nuts and bolts and understand every single thing that's



going on in the car in order to be able to use it effectively. I think that kind of reiterates your point about data science, that that is exactly where the world is going, because this field is growing so rapidly and people and companies and the world just needs more skilled people who can apply data science and that creates a huge world of potential and opportunity for those who want to break into this field.

Dan: Yeah.

Kirill: Wonderful. Thank you so much for sharing that and for coming on the show. If our listeners want to contact you or find you or follow your career and see what else you're going to get up to in the future, where is the best way to connect with you?

Dan: I have a website, it's danshiebler.com, that should have my contact info on it. I post some interesting things that I'm thinking about and working on there. My e-mail address is danshiebler@gmail.com or dan@gotruemotion.com. I respond to e-mails from people who have questions if anybody is interested in some of the stuff that I'm working on.

Kirill: It's definitely great that you've shared you're happy to respond to questions. And is it okay for people to connect with you on LinkedIn?

Dan: Absolutely.

Kirill: Fantastic. And one more question I have for you: What is your one favourite book that you can recommend to our listeners so that they can become better data scientists?

- Dan: It's not really a book, it's a collection of online course notes, that sort of form a textbook. It's Andrew Ng's CS 229 course that he teaches at Stanford. The course notes of the course are all online. If you just google CS 229, it's the first thing that comes up. And it's really an incredible dive into some of the theory behind machine learning and data science with incredibly intuitive description. So, for somebody who wants to get a little bit deeper in there and feels relatively comfortable with multivariable calculus and linear algebra and such, it's the tool that I use to really introduce myself in the cutting edge in the field. I find myself going back to it again and again to refresh on different things. I highly recommend it.
- Kirill: Gotcha. Thank you so much again. So that is Andrew Ng's course notes for CS 229. Thank you very much, Dan, for coming on the show, taking out some time of your Sunday evening to share all of these insights. I'm sure a lot of people will find them very useful.
- Dan: No problem. It was great being here. Thank you for having me.
- Kirill: So there you have it. That was Dan Shiebler from TrueMotion. I hope you enjoyed today's podcast and learned quite a lot of new things. Of course, there were so many valuable things that Dan shared with us. For instance, all the sensory data from smartphones, how it's collected and processed, how they analyse the orientation of the phone, that was so interesting, and how they apply data science to extract valuable insights of our people's driving. Also it was very interesting to learn about Dan's background.



But probably for me, the one most fascinating thing and the most inspiring thing from this podcast was how Dan in his spare time researches other areas of data science that he has interest in. He was talking about how he does research in the space of convolutional neural networks in his spare time just because he is very interested in that field. That is a huge testament to human ambition. If you really want to get into data science, if you really want to excel at something, then you don't have to just wait until you get a job or you don't have to only learn about the things that you're doing at your job. As Dan showed us with his example, you can be investigating and learning about completely different fields of data science if you have the passion and drive to do so.

Hopefully that inspired you as much as it inspired me. And remember to hit up Dan and connect with him on LinkedIn and check out his website, danshiebler.com. We'll definitely put all the links and resources from this episode into the show notes, which you can find at www.superdatascience.com/59. And if you enjoy this podcast, then make sure to go to iTunes and leave us a rating. We would really appreciate that to help spread the word about data science into the world. I can't wait to see you next time. Until then, happy analyzing.