

Sydney Hernandez

SSAC2020

In November of 2019, I got my ticket for the 2020 MIT Sloan Sports Analytics Conference. I honestly could not believe that I was actually going to go. As they continued to post the speakers, my excitement continued to grow. When they released the schedule, I printed out several copies. I started picking out which panels I was going to go to. At the beginning, I only had 3 time slots where I had picked out one panel to go to. The rest of them I could not decide on. I went to my teachers for their advice on which sessions I should go to. It was tough, but in the end, I'm glad I picked the ones I did.

Day 1

The first day of the conference was Friday, March 5, 2020. I was SO excited! The first panel I went to was, "Analyzing Film and Data Concurrently to Get an Upper Hand", and the Texas A&M Football Director of Sports Science and Video Coding was the speaker. In that session, I learned that coaches and analysts don't pay attention to just the data the fans see. They pay attention to other metrics, like the players' heart rates. There was one player a few years back, who had great speed, could catch and throw, but always messed up the plays. When looking at his heart rate throughout the practices, they realized that it always spiked right before certain plays. They looked at other data and realized it spiked because he was nervous because he didn't know the play. I learned that they do a SWOT analysis for every player, team, and potential prospect. They use this and prospects' stats to determine if they fit into the roster. He made it clear several times that they do not take a player just because their amazing stats, but because there is a hole in the roster that that player can fill and fill well.

Next, I went to "Football and CTE: How New Data Can Save the NFL". I learned about how delaying the age at which kids begin playing tackle football drastically changes their chance of getting/having CTE and a career in professional football. I learned that for every 2.6 years a person plays tackle football, their chances of CTE double. If kids started playing tackle football at age 14 (high school) rather than age 5, they are ten times less likely to get CTE. The age kids begin to play tackle football at needs to be pushed back for their safety. However, it seems as though kids are getting recruited younger and younger every year, which means they have to get into the game as soon as they can. This is not just an issue in football, I see it in the athletes at my school. I see it with my brother. Kids feel the pressure to be great and commit or get drafted

as soon as they can. They all want to be the freshman that makes varsity and be a ranked player. Most kids cannot just start playing tackle football in high school and then sign with a D1 school or get drafted. Pushing back the age kids begin playing tackle football would also probably result in less future NFL fans. When kids play for 10-13 years, they have a connection to the game and are more likely to watch the NFL and be fans. In summary, if youth football is not reformed, CTE will not improve, and the NFL will suffer.

The third session I went to was, "Hockey Analytics: A Faceoff with Data". In this panel, I mostly learned about how they could track more data, how they use data, etc. Location tracking data was really big and was a topic for discussion in nearly every panel/session. They talked about collection location tracking data by using a new puck with a chip in it and making the players wear tags. The NBA and Soccer leagues use similar data because of the size of the ball and the speed. All of this data will show what makes an effective team, player, or plays. In the last 7-9 years, there has been a trend towards speed in hockey. This is another metric location tracking data is useful for gathering. There is no doubt about it; data analytics has changed hockey's in-game strategy. With all of this access to data, hockey is more measurable, data now drives competitive strategies, and they develop metrics about first plays (controlled vs uncontrolled). They also now develop strategies, then test, measure, and evaluate these strategies. All panel members agreed, communication of these results is key. You, a data analyst, might see the trend or the issue in a play or player, but if you cannot effectively communicate this to the coaches or players, your findings are pointless. The whole point of this side of sports data analytics is to look at the data and find what is wrong or could be better and pass it along to the players and/or coaches, but if they cannot understand what you are giving them, then the problem cannot be fixed and there will be no improvement. In the end, it all comes down to communication!

I loved this 4th session! It made me so excited for the future MLB At Bat app that we will (hopefully) have access to soon! This panel was called, "Innovating for the Next Generation of America's Pastime". In case you did not know, MLB recently partnered with Google Cloud and now all of MLB's data (player performance, app, video data, fan data, etc.) is being stored on one platform with Google Cloud. This is HUGE, and soon fans will be able to see the amazing results this partnership brings. I am a HUGE Chicago Cubs fan, and also a really big Willson Contreras fan. I am also a high school student and do not have the time to watch every play of

every single one of the 162 games of the regular season (even though I wish I could). In the near future, I will be able to go into my MLB At Bat app and look up Willson Contreras' 0-2 hits or Javy Baez's defensive plays. I can choose to see all of them from their entire career, this season, this week, this game, or in the 4th inning. You can be as broad or as specific as you want. You can see when a squirrel, a bird, or cat was on the field. From the app, you will be able to share this short video to social media. Now all of these features will not be available all at once, but we can expect to see this in the next 2-5 years! The video highlights, known as "Project Fastball" internally, will be the first phase of this big innovation. Today's children cannot sit still and watch a 2-3-hour baseball game. Everyone can and will admit baseball is a slow game, but it is a great game. In order for MLB to have a substantial fanbase in the future, they need to improve fan engagement. That is what this video snippet feature will do. Because fans can slice and dice video of the game, they are still engaged as a fan, even without watching the 3-hour game. Baseball fan engagement has also increased since offering an exclusive game of the week only available on YouTube. MLB has also start "reference bottling" and targeting the youth via YouTube. They then apply it to the ballpark. There was discussion of AR in ballparks and CLEAR ballpark bartenders. Basically, you will go and get whatever beer you want, and sit it on the mat, and it will know exactly what the product and charge is. Then, you will use your biometrics through CLEAR to verify you are at least 21 years old and charge your credit card. This was available at Citi Field last season and the feedback was great, and more ballparks are likely to adopt this service in future seasons. Next, they talked about how access to all of this data has changed the game and resulted in new skills. This topic also makes me really excited for reasons I will elaborate on later. In the recent years, there has been a trend in exit velocity and launch angle. Players have been using this data to train their swing to produce a hit with a certain launch angle. Going back to 2015, there has been an increase in launch angle each year. This is a result of data being available and analyzed, then communicated to the players and coaches. As more and more data becomes available, new skills are going to come up. An example of this is pitch framing. Pitch framing has become an important skill for catchers to learn since data and video analytics has increased in popularity. One thing I will point out is that the use of robotic umpires will mean that pitch framing will no longer be a necessary skill catchers need in the Major League.

The next two I went to were very similar, and were actually focused on the programming side of data analytics with location tracking and Google Cloud. Using Google Cloud, they want to take sports data analytics to “the next level”. This means making player tracking and data collection faster and easier at a new, larger scale. To make things easier, they will use Big Query ML linear and logistic regression models, k-means clustering methods, and matrix factorization. They will then enable new features like Cloud video intelligence, video intelligence API. They can then import video data into Big Query to get insights.

My seventh session of the day was, “At the Intersection of AI & Human Potential”. Dafna Aaronson, founder of Be Sports Minded LLC., was the speaker and was more focused on the intangibles. She focuses on the if, why, and how players will be successful in a given environment. Her machine learning platform identifies these intangibles without ever testing a player. This machine learning is useful for scouts, as it gives them a leg up. She recognized there is a problem. Unreliable psychometrics, false positives, and systematic inefficiencies. When a player is asked questions, they know what they should say and it is not always their honest answer. This is a false positive. Unreliable psychometrics comes from relying on self-reporting, which lack internal validity and is biased. Systemic inefficiencies often come with different skills and scouts, making sure everyone is on the same page of what you are looking for, defining character traits, etc. She uses natural language and IBM software to find anything on the internet an athlete has said or written (not about the athletes by someone else). This is then analyzed to search for 5 standards that are stable by age of 12. You can then look at younger athletes to see if they have what it takes to handle adversity and keep motivated. In summary, her company use advanced analytics to improve the reliability of predictive modeling and risk analysis and protects franchise or prospect investments.

My dad has worked in the parking industry for 18 years, and the one session he told me I had to go to was, “Data Driven: How the Giants are Using Analytics to Satisfy Parking Demands”. In 2018-2019, Oracle Park had 2,500 parking spaces. The ball park holds 41, 915 people. While public transportation is big in San Francisco, it isn’t big enough to wear you only need 2,500 parking spots. This year, in 2020, they will begin phase 1 of their mission to solve this parking issue. However, they will only have 1,800 parking spaces the 2020 season. They have come up with plenty of options for how to deal with their parking situation, but with the all of the rumors regarding the 2020 season, the Giants might not have to worry about this.

The last panel I went to was, “Making the Right Play: Optimizing the NBA’s Season Structure”. Currently, the NBA season is 82 games. Steve Koonin, CEO of the Atlanta Hawks, emphasized several times that relevance equal revenue! If you create relevance the revenue will fix itself. By condensing the season, every game means more, but you lose rest days for veterans or practice days for young teams. He also had the idea of starting the week after Thanksgiving, “let football have its time”, and (I thought this was kind of funny but true) “If King Kong is at your door, you might go out the back door, rather than go out the front and engage in a hand-to-hand fight with King Kong. Many times, at the start of the NBA season, we are competing with arguably the best Thursday Night Football game with the NBA on TNT, our marquee broadcast, and we get crushed and we wonder why”. I learned that you have to plan around your barriers (in the NBA’s case football), don’t fight it, just adjust. You also have to create relevance and purpose or you’re never going to generate revenue.

Day 2

I was SO excited for this panel, “Moneyball to Homerball”, with THE Bill James and Tom Tango. I learned so much and was smiling for the entirety of the panel. I was told to do what nobody is doing, nobody has to tell you what to do, just do it. There are no limits, so keep searching and looking. I learned that just because everybody knows about it it doesn’t mean it isn’t a competitive advantage. Do not exclude data. If you need ideas of what to research, look at other’s work, then expand on it. You have to have a question. What are you trying to answer with the data? Roles are just going to keep adding and going. The more general your degree, the more it will hurt your opportunity to join a club. You have to have something to offer nobody else has. Drill down on the smaller, more narrow questions, but make sure to still focus on the bigger questions. I know that all of that was really choppy, but those were some of the little lessons or things I learned for getting your career in sports analytics started. Besides metrics, they talked about robot umpers, increasing pace of play, etc. They talked about moving the plate up, dropping the mound 2 inches, and moving the plate back 21 inches. I thought their discussion about leverage index was really cool. It is all about determining how crucial a play is. When you really think about it, the emphasis on closers is way more than their true value. Your closer is more likely to go in in the 9th if you have a 3 run lead than if it was a tied ballgame. If people understood leverage index, the roster might look different. Pitch framing didn’t matter 15 years ago, but research magnified it. The shift has resulted in less balls in play. This makes pitchers

look better and easier to determine the best pitcher. They debated a 4-man rotation vs a 6-man rotation. They pointed out how illogical it is to increase the number of days between starts of your best pitchers. It all starts with one team below .500 that is desperate for a solution, like the Rays in the mid-late 2000s.

The second session of the day was, “Intangibles: Unlocking the Science and Soul of Team Chemistry”. I thought this one was pretty cool because I could see what she was saying with my baseball team. Team chemistry IS important. Everything about a person is contagious. She had this story about the Red Sox in the World Series and this last-minute lineup change because when this one guy, Johnny Gnomes, played, everyone seemed to play better because they were excited to play with him. It reminds me of this one time when our starting pitcher ran and hugged our catcher because he was that excited to play with him. They looked just like the picture on the right. These baseball players are together so so much and there is no



escaping the influence others have. It’s all in their head, but belief can elevate performance. Team chemistry does not win the game, but it does lead to better individual and team performance. Sometimes, there is task trust. They only trust they will do their job and care as much about winning as they do. You still need team chemistry, you need bonding, you need that sense of completeness. “The new advantage will go to those who figure out how to recognize it, foster, and harness the power of this hidden nation-state we call team chemistry.” (Joan Ryan) Get rid of your super-disputers, your complainers, because they will not change.

I thought the 3rd session was also really cool because it covered what I had talked to my sports marketing teacher about. “The Power of Location Data” was about tracking phone signals before, during, and after sports events to see what they do, where they go (within a radius of the venue), the times they get there, leave, etc. They use this data to connect their brand to the customer’s lifestyle. It can also be used to predict fan behavior, personalize their experience, raise the value of sponsorships, improve broadcasting and advertising, etc. It was all about using fan location data and turning it into how they would reach and sell to them.

I went to a few other panels before the women’s lunch-in. One (Defending the Pass) focused on quantifying defensive passes, holding them at the line, etc. The other was about

learning across multiple sports. They talked about baseball could translate to soccer, football to basketball, etc. Bringing in people from different sports or industries can be really refreshing, you just have to go for it and be aggressive. You cannot afford to be afraid to try something that seems crazy, you have to keep trying until something works. Make the two worlds alike. If you are a physics major, you can understand arc shot or launch angle. It becomes so much easier to learn another sport once you connect what is similar.

Next I went to the women's lunch-in and met so many really cool people. My table lead was awesome and is Manager of Strategy and Analytics at the Baltimore Orioles. I met some students from UMass, University of Maryland, MIT, and Princeton. I met worked for NASCAR or were entrepreneurs... one of them started her own company and later hired her old professor! I was amazed by these women and we talked for hours. I accidentally missed one of the sessions I had planned on going because I was talking to these really awesome people. When I wasn't in a session or at the lunch-in I went around and talked to people and talked to a bunch of the start-up companies. I gave out a ton of the business cards I had made. Most of the people there did not believe I was in high school and they couldn't believe it. The fact that they thought I fit in and I did not feel out of place there makes me so happy and excited!

The last session I went to was also one of my favorites. It was funny and I realized how incredibly lucky I am. The panel, "Sports in 2040: Hindsight is 2020", featured Daryl Morey, Jessica Gelman, THE Bill James, Nate Silver, and was moderated by Katie Nolan. They started with what they thought the in-game experience would be like. Most of them said that it would be much quicker to get in and out of the ballpark. James made a really funny comment about how everybody is going to know who their sitting around and everything about them. Morey and Silver talked a bit about in-game, real time sports gambling (play by play). Bill James and Daryl Morey made a great point about how sports and games always need to keep uncertainty in the outcome of the game. I had never thought about it like they were from the stadium perspective. If the home team is winning and the score is not even close, people will begin to leave and then you miss out on money from concessions, merchandise, etc. When it came it making baseball faster... there were some interesting ideas. There was the expected pitch clock idea, but then there was moving the mound back and the batter not being able to call time. Daryl Morey cuts in and says that you have a robot ump and the pitcher can just keep pitching and if the batter isn't at the plate then too bad! They that ball or strike. The way that he explained it was hilarious. They

then talked about women in tackle football, robot umps, circular basketball courts, and so much more. I can't wait to see what actually does happen by 2040 to see if or how many of their predictions were true.

Mr. O,

THANK YOU! I learned so much more than I ever thought I could and met so many people. Thank you for giving me the opportunity to explore sports analytics, I loved it! I could not have asked for a better experience! I seriously could never thank you enough. I know this was really long, but I really did learn so much and I feel like I didn't even cover half of it!

Sincerely,

Sydney Hernandez