

Big Data Analytics: Homework 1
Dan Tully
Syracuse University

Big Data Analytics: Homework 1

Big data can be found in many businesses and industries, from healthcare treatment options to automobile performance sensors, these organizations gather data points all the time which impact operations and decision making. Another organization is the National Football League (NFL). I chose the NFL because with the data gathered from the Players, the sports analysts and coaches use it to predict the outcome of plays and ultimately the game.

How do they use this big data?

They call it the “NFL Next Gen Stats [(NGS)] — developed in partnership with Zebra Technologies, Wilson Sporting Goods and running entirely on Amazon Web Services (AWS) infrastructure — [which] provide clubs with data to analyze trends and player performance, while enhancing the fans’ experience in-stadium, online and during game telecasts... [They use] 2–3 radio-frequency identification (RFID) tags installed into the players’ shoulder pads...20–30 ultra-wide band receivers...RFID tags on officials, pylons, sticks, chains, and in the ball...[with] an estimated 250 devices are in a venue for any given game...The raw data is used to automate player participation reports, calculate performance metrics, and derive advanced statistics through machine learning (ML) on AWS. More than 200 new data points are created on every play of every game.” (NFL next Gen Stats | NFL Football Operations, n.d.) Figuring the “NFL Average Number Of Plays ...[around]... 63.24 plays per game played in 2021.” (NFL Average Number of Plays per Game 2021, n.d.) Also, in 2021, was the first time they introduced the “18-week, 272-game regular season schedule” (2021 NFL Schedule Announced | NFL Football Operations, 2021). Doing the math; it looks like the data points collected on every play (200) multiplied by the average number of plays per game (63.24) multiplied by the number of regular season games played (272) gives us 3,440,256 data points from just the RFID tags during the regular season in 2021.

As you can imagine, player performance is one of the main goals of data collection in any sport and the NFL is no different. Unfortunately, any number of ...”features — down and distance, distance to the goal line, elapsed game time, length of the current drive, temperature — could, in principle, affect player performance.” (How AWS Scientists Help Create the NFL’s next Gen Stats, 2021) There are a lot of variables to consider, some are in addition to what these RFID tags can provide. For example, the weather and game clock data is not available through the RFID tags but can still be captured and sinked with the data that is collected from the tags to fill in some additional variables to help with performance predictions. The NFL may want to use both the RFID tags to analyze a runner's speed and weather data to capture the temperature. They may discover that a runner is extremely fast in warmer weather but notably slower in colder weather or that extreme weather has minimal impact on the runner's performance.

Player safety is another one of the uses for the data. “The Digital Athlete, a joint effort between the NFL and AWS that represents the next generation of player health and safety for the league...The Digital Athlete uses artificial intelligence (AI) and ML to build a complete view of players' experience, which enables NFL teams to understand precisely what individual players need to stay healthy, recover quickly, and perform at their best. In time, the technology may be able to help predict and prevent injuries.” (Langton, 2024) One way the NFL can prevent injuries is by using the data they capture and evaluate the data around any injuries caused. They can take that information to retrain players with the new football techniques, implement new

league rules, and improve player equipment in order to prevent or mitigate the risks of future injuries.

Fan user experience (UX) is yet another use for the data. “At key points, the broadcast [is] augmented by real-time evaluations using the NFL’s NGS powered by AWS. Several of those stats, such as pass completion probability or expected yards after catch, use ML models to analyze the data streaming in from RFID tags on players’ shoulder pads and on the ball” (How AWS Scientists Help Create the NFL’s next Gen Stats, 2021) This has really made watching the games more fun for fans by providing probability data and relevant statistics on players while providing coaches with additional realtime information to aid them with calling plays.

What are the ethical dimensions with collecting this big data?

As the NFL NGS collects over 3.5M data points through the regular season and postseason, we want to consider some of the potential ethical dilemmas associated with NGS. The NFL certainly wants to protect themselves from lawsuits but also wants to get a return on their investments with the RFID tags. One article implies that the NFL gathering data since 2017 didn’t know what to do with the data so they turned to AWS to help develop ML algorithms and stats which can be used for marketing purposes. (Lies, Damn Lies, and NFL next Gen Stats Powered by AWS · Skewed, n.d.)

I don’t think the new data specifically gathered by NGS is any more intrusive for the Players than what they have already been exposed to by the NFL with manual observation and video analysis. “150 Million feet of footage (and counting) ... The legendary NFL Films vault is the largest archive of sports film and video in the world.” (NFL.com | Official Site of the National Football League, n.d.) Players probably consent to this new tracking as they have consented to video taping by the NFL for decades directly in their signed NFL contract deals.

Data accuracy is important for the NFL and AWS. They have “an estimated 250 devices are in a venue for any given game. A team of three operators is required at every game to confirm that all tracking systems are functioning properly.” (NFL next Gen Stats | NFL Football Operations, n.d.) Since they have many people relying on these states it is probably very important to legitimize them. As we discussed earlier they are using this data for safety protocols, player performance (which I’m sure includes trade decisions) as well as the UX for the fans to follow all of the action on the NFL app and/or website.

While considering transparency, bias, and fairness of the data we know that NGS provides easier data collection over the previous manual methods and that the NFL has teams explicitly working to ensure the equipment is functioning properly. Bias does exist in the data since racial stereotypes inherently still exist in drafting players. (Delucchi, 2023) “The [NFL] has started sponsoring Kaggle competitions in an effort to crowdsource useful player-tracking stats, albeit with carefully controlled samples of data under restrictive terms” (Lies, Damn Lies, and NFL next Gen Stats Powered by AWS · Skewed, n.d.) The NFL also has an annual Big Data Bowl competition “to contribute to the NFL’s continuing evolution of the use of advanced analytics. The crowd-sourced competition uses data and technology to spur innovation that results in creating new insights, making the game more exciting for fans and protecting players from unnecessary risk.” (NFL Big Data Bowl | NFL Football Operations, n.d.)

Wrap up!

As the NFL NGS continues to collect millions of data points annually from their games and Players, we can see how collection and storage of this data could have been an issue. Partnering with AWS seems like the perfect solution for mitigating any of their data architecture problems. I believe this newer NGS technology does really enhance the UX for the fans while making it easier for sports analysts and coaches to review player performance. It can also lead to player satisfaction by reducing injuries and extending players careers. This research, in my opinion, has not surfaced any new ethical dimensions not previously known about in the NFL.

References

- NFL Next Gen Stats | NFL Football Operations. (n.d.). Operations.nfl.com.
<https://operations.nfl.com/gameday/technology/nfl-next-gen-stats/>
- Langton, J. (2024, February 1). Building a digital athlete: Using AI to rewrite the playbook on NFL player safety. NFL.com.
<https://www.nfl.com/playerhealthandsafety/equipment-and-innovation/aws-partnership/building-a-digital-athlete-using-ai-to-rewrite-the-playbook-on-nfl-player-safety>
- How AWS scientists help create the NFL's Next Gen Stats.* (2021, February 5). Amazon Science.
<https://www.amazon.science/latest-news/how-aws-scientists-help-create-the-nfls-next-gen-stats>
- NFL Average Number Of Plays Per Game 2021.* (n.d.). StatMuse. Retrieved April 28, 2024, from
<https://www.statmuse.com/nfl/ask/nfl-average-number-of-plays-per-game-2021>
- 2021 NFL Schedule Announced | NFL Football Operations.* (2021, May 13). Operations.nfl.com.
<https://operations.nfl.com/updates/the-game/2021-nfl-schedule-announced/>
- Lies, Damn Lies, and NFL Next Gen Stats Powered by AWS · Skewed.* (n.d.). Skeweddata.github.io.
https://skeweddata.github.io/2021/01/24/nfl_stats_lies.html
- NFL.com | Official Site of the National Football League.* (n.d.). NFL.com. Retrieved April 28, 2024, from <https://www.nflfilms.com/licensing/>
- Delucchi, M. (2023, September 11). *New evidence shows how NFL discriminates against Black QBs.* SFGATE.
<https://www.sfgate.com/49ers/article/nfl-draft-analysis-racial-bias-quarterbacks-18355172.php>
- NFL Big Data Bowl | NFL Football Operations.* (n.d.). Operations.nfl.com.
<https://operations.nfl.com/gameday/analytics/big-data-bowl>