

Developing State-Based Recommendation Systems for Golf Training
(Technical Report)

Sports Analytics: An ethical assessment of current regulations, current sources of risk, and future sources of risk
(STS Topic)

A Thesis Prospectus in STS 4500
Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia
In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science in Systems and Information Engineering

Author
Alanna Flores
October 30, 2019

Technical Project Team Members
Kelly Rohrer
Jacob Ziller
Christopher Kaylor
Orlando Jimenez

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Signature _____ Date _____

Approved _____ Date _____

Rider W. Foley, Department of Engineering and Society

Sports Analytics: An ethical assessment of current regulations, current sources of risk, and future sources of risk

Introduction

Sports analytics, the concept of using data informed decisions to achieve an upper hand in performance, has become increasingly popular ever since the book “Moneyball” by Michael Lewis (2013) transformed the baseball industry. The latest innovations in sports analytics are presented at the MIT Sloan Sports Analytics Conference every year, where baseball, basketball, and football have a dominating presence because of their aggressive induction of sports analytics (Sweeney, 2019). Despite the prolific popularity of golf, an approximately 24 million golf players in the US, golf has only increased the usage of analytics in the last few years (Gough, 2019). Analytics companies like ShotLink and ShotByShot currently offer advance to the Professional Golfers’ Association, but the game is not close to achieving the same level of implementation as other major sports (Dusek, 2018). Beyond the PGA stats and analysis, amateur and collegiate players don’t have technology to rely on when working to improve their score in an efficient manner.

This is the problem that our technical team is working on to make possible, using a small subset of round statistics from a player and providing a recommendation for drills that will effectively improve their next round of golf. The advancement into data-driven approaches can be seen trickling down into all sports. A research report expected growth of “the global sports analytics market... [will reach] a revenue of \$4.5 billion by 2024” (Infoholic Research LLP, 2018). For data analysts, this expansion is an opportunity to find new breakthroughs hidden in athletic data, but many stakeholders in the global sports market have not considered all of the implications that could result from this expanse of sports analytics.

As sports analytics grows at an exceptional rate, concerns about the negative implications of overusing the data without adequate regulation grow in tandem. While working to achieve discoveries in our technical project, our team has felt the eagerness that comes from new discoveries. This eagerness and push for more valuable findings can cloud judgement on what advancements will implicate athletes in the

future. Responsible engineering comes when all outcomes, negative and positive, are addressed. Tackling the negative repercussions for athletes should come first before the advancements in sports analytics are employed.

Technical Topic

Golf is one sport that recently began to analyze massive amounts of statistics to inform decisions regarding play, training, and player evaluation. This recent push into data-driven analysis has made companies like Game Forge, trailblazers in the field of golf analytics. Beforehand golfers, amateur and professional alike, would rely on age-old knowledge and training methods that may have no statistical significance in effecting play. The golf market is being flooded with “swing analyzers,” “smart golf clubs,” and “GPS based stat tracking tools” (Digest, 2019), all of which are using stats to inform a players’ decision. What’s not yet developed is a program that recommends a training regimen that could use a concise set of input statistics to effectively improve a player’s score.

Game Forge has enlisted our technical team to use their extensive and consistently growing data from PGA tour, LPGA tour, collegiate, and amateur players. Using the provided data we will formulate a data-driven training recommendation system. A player will input roughly 20 statistics per round, relative to par score, birdies, bogies, etc. From these performance metrics we will create a methodology to form a recommendation on how to improve score. This process takes into account the effect of the Game Forge drills on performance, past performance of a golfer, and golf course difficulty. Our technical team aims to provide our client with a general training regimen process and an in-tournament training recommendation system to fully encompass the purpose of this technology.

Half of the team has focused on continuing and refining this method by using the updated dataset of round scores to cluster players by score. For example, one cluster may be made up of golfers that scored a 2 relative to par and performed well on the greens. Another cluster may be made up of golfers that score a 2 relative to par and performed poorly on the greens but made up for it elsewhere. The other

half of the team has focused on working with the drill data, a dataset that includes performance on drills, among other variables. This subgroup is working to answer questions such as: Does score on a particular drill predict round score? Over what period of time should drills be considered when used to predict round score? Should drills performed long ago be weighted less than drills performed yesterday? If so, what is the function governing this decay? The team's current model is a simple linear model predicting round performance based on drill performance. Moving forward, the team will look to improve the model by incorporating clustering and time series analysis to better predict round performance based on drill performance. From here the team will be able to identify areas for improvement for individual players. These areas will likely be the game categories or skill categories defined by Game Forge in which players should practice drills.

This drill recommendation system provides an easily accessible tool that would provide valuable insight with minimal effort from the players. Since most golf analytics companies require 500 statistics per round, Game Forge's 20 statistics are met with much relief from players. The value of sports analytics can provide a competitive advantage for players, but in other cases problems can arise when analytics on player data go too far without proper consideration for the implications.

The Need for Ethically Responsible Sports Analysis

The development of a drill recommendation system for golfers will provide players with training techniques to improve play without negative side effect, but when high-powered data analytics starts using biometric and performance data to value players, ethical and legal problems start to arise. Analysis of athletic data can provide insights into improving performance, mitigating injury, and streamline the recruitment process, but with those advancements comes some major threats for the athlete's themselves. Current regulations are insufficient when it comes to the collection and usage of biometric and performance data from professional and collegiate athletes, that leaves players exposed to "risks of exploitation, coercion, and employee discrimination" (Karkazis & Fishman, 2017). To maintain a

responsible innovation for the society directly affected, the athletes themselves and their interests, primitive regulations must be enacted to prevent the unintended consequences.

Anticipatory governance framework focuses on creating “future oriented dimensions of responsibility... that offer greater potential to accommodate uncertainty and allow reflection on purposes and values” (Stilgoe, Owen, & Macnaghten, 2013, p. 1569). This urges policy-makers and regulators to work with data analysts to foresee the future developments and ensure all athletes are protected from the negative implications that could arise. New innovations in sports analytics must be met with social and political responsibility on the other side if it is going to be of any help to our society in the future. The current regulations and laws in place that address sports analytics need to be modified and adapted to each new development and concern discovered. For collegiate athletes that aren’t protected under Health Insurance Portability and Accountability Act, “NCAA does not currently address the use or collection of performance data in its bylaws” (Smolenski, 2019). This leaves athletes’ data to be used at the discretion of each specific team. Even as professional athletes, the protections in place are no more regulated or clear since “biometric and biomechanical data are typically not categorized as personal health information under existing federal framework” (Osborne, 2017). This exemption in federal law can be used by teams as an excuse from complying with federal requirements. These nonspecific and unsupportive regulations leaves the collection and use of athletic biometric data up to team preference, which in most cases isn’t in the best interest of the athletes.

In the article “Beyond Regulation: Risk Pricing and Responsible Innovation,” the idea of innovation governance is described as “the need to drive responsibility ‘upstream’ in the innovation process” which is precisely what needs to happen in the world of sports analytics (Owen et al., 2009). The ethical responsibility to players needs to be prioritized among stakeholders in initial phases of the technologies rather than post implementation, consequential issues, then eventual policy enactment. Reconstructing the framework of developing sports analytics responsibly will have to consider the multitude of stakeholders and different categories of data collected, biometric and performance amongst

collegiate and professional. The actors in charge of creating the regulations are team managers, league policy makers, lawmakers, and the engineers, and data scientists. Each stakeholder has some ability to regulate the negative outcomes with preventative measures. This regulation could occur in the form of contracts, security measures, legal bindings, league wide by-laws, and more.

Opposing stakeholder perspectives bring challenges between maintaining the ability to allow improvement in sports analytics to continue, while insuring the improvements don't cause an ethical issue. On a professional and collegiate level athletes are facing the decision to release their data for analysis without comprehensive knowledge of the future outcomes or risks. The athlete's biometrics, "the measurement and analysis of any particular physical characteristic", or health data can include sensitive information about a player (Osborne, 2017). On the other hand the performance data, game or practice statistics, are becoming so honed and advanced with analytics that even the slightest measurement could mean the difference between receiving a scholarship or not. Coaches and managers of teams are implementing the collection of both types of data, often forgetting to respect the rights to privacy of a player. The sources of risk in implementing analytics include threats to privacy, ownership of data, misinterpretation of data, confidentiality, data security, and coercion of players. Even though athlete's careers are based off of their performance their need for protection from these risks is still imperative.

Research Question and Methods

This research aims to address: What are the unintended consequences of utilizing sports analytics without sufficient regulation? Analyzing how the athletes are at risk and where there is cause for new regulation or governance will aid analysts, players, team managers, and leagues to implement sports analytics in a responsible way. The research will offer case studies of how analytics have been misused or created negative consequences for athletes, and aims to identify the sources of risk. The case studies the dimensions of responsible innovation to highlight the ethical issues at hand. Case law will be analyzed for its content to realize consequential ambiguity and necessity for revisement. Using policy as evidence with content analysis from case studies the research will address the areas of concern and where the current

regulations fail. This analysis will be accompanied by first hand interviews with athletes about their experience and opinion of the data collection and usage in their field. In addition to the research on current issues, possible future developments in sports analytics will be taken through the same ethical assessment to address future concerns through the anticipatory governance framework. This method of research takes into account the perspective of athletes from the case studies and synthesizes the requirements that are not being met, which will give a comprehensive view of the situation.

Conclusion

At the conclusion of our work, our technical team aims to deliver Game Forge a complete drill recommendation system built off of statistical evidence. This simple-to-use system will provide golfers with easy access to well-proven training methods that will hopefully improve their game. The ethical implications research will provide a comprehensive ethical assessment of current regulations, current sources of risk, and future sources of risk that result from the implementation of sports analytics. The resulting scholarly article will aid in the transition to data-driven sports that also mitigates unnecessary risk for players with an outline of areas of demand for responsible regulation and innovation.

Timetable of work to be carried out

November 16th: Returned prospectus feedback (peer review and instructor review)

November 26th: Signed prospectus due

By December 6th: Case studies identify sources of risk, ethical assessment and content analysis

By December 20th: Case law assessment and content analysis

December 14th-23th: Organize Interviews

January 3th-10th: Research future developments and take through ethical assessment framework

January 10st-15th: Synthesize research

January 15th-February 15th: Write STS Thesis paper

February 15th-March 6th: Revise

March 7th- March 30th: Finalize

References

Dusek, D. (2018, March 5). By the numbers: Analytics become an increasingly important part of golf.

Retrieved from <https://golfweek.com/2018/03/04/by-the-numbers-analytics-become-an-increasingly-important-part-of-golf/>.

Golf Digest. (2019, April 23). Best Swing & Game Analyzers. Retrieved from

<https://www.golfdigest.com/story/best-golf-swing-and-game-analyzers>.

Gough, C. (2019, August 23). Golf: number of participants U.S. 2019. Retrieved from

<https://www.statista.com/statistics/191907/participants-in-golf-in-the-us-since-2006/>.

Infoholic Research LLP. (2018, November). Sports Analytics Market Forecasts up to 2024 - Research and Markets. Retrieved October 30, 2019, from

https://www.researchandmarkets.com/research/wlpbw8/global_4_5?w=4.

Karkazis, K., & Fishman, J. R. (2017). Tracking U.S. Professional Athletes: The Ethics of Biometric Technologies. *American Journal of Bioethics*, 17(1), 45 - 60.

Lewis, M. (2013). *Moneyball: the art of winning an unfair game*. W.W. Norton. New York.

Osborne, B. (2017). Legal and Ethical Implications of Athletes' Biometric Data Collection in Professional Sport. *Marquette Sports Law Review*, 28(1), 37–84. (s3h:128155321).

Owen, R., Baxter, D., Maynard, T., & Depledge, M. (2009). Beyond Regulation: Risk Pricing and Responsible Innovation. *Environmental Science & Technology*, 43(18), 6902 - 6906.

Smolenski, G. (2019). When the Collection of Biometric and Performance Data on College Athletes Goes Too Far. *Wake Forest Law Review*, 54(1), 279–301.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a Framework for Responsible Innovation. *Research Policy*, 42(9), 1568 - 1580.

Sweeney, M. (2019, March). AimPoint Golf – Accelerated Skill Acquisition in Putting. Retrieved from <http://www.sloansportsconference.com/content/aimpoint-golf-accelerated-skill-aquisition-in-putting/>.