Thesis Project Portfolio

Golf and GameForge: Innovative Analytics for Recommender Systems

(technical research project in Systems Engineering)

Predictive Analytics in Sports: Offsetting Human Bias

(sociotechnical research project)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Thomas Lee Twomey
Spring, 2022
Department of Systems Engineering

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.

Thomas Twomey

Technical advisor: William T. Scherer, Department of Systems Engineering

STS advisor: S. Travis Elliott, Department of Engineering and Society

Table of Contents

Background

Winning in sports stems from the success of the great leadership, a robust front office, and a superior coaching staff. In the past, decisions about which players to draft, trade, and play were a product of the gut feeling of managing staff, as opposed to an objective, quantifiable method. This practice was forever changed in all of sports when Billy Beane, General Manager of the Oakland Athletics baseball team, used statistical analysis to discover the secrets of success in the imperfect science of evaluating baseball players in 1997 (Steinberg, 2015). This was the first known use of prioritizing data and statistics to drive decision making in all of professional sports. This story was the inspiration of Michael Lewis' famous book *Moneyball*. Since the "moneyball" approach was first documented, sport analytics has grown into the large focus that it is today. As a result, the success of professional athletes is rarely reported without relevant numbers and statistics.

General Research Problem

In the U.S., how have sports organizations used analysis to promote fairness?

Ostensibly, the rules of athletic competition give all competitors an equal chance at success. In reality, most sports are far from a level playing field. Analytics, such as evaluating a player's batting average or on base percentage, offer quantitative evaluations that may supplement or substitute for qualitative evaluations, like a player's physical size, strength, and appearance. "Fairness is part of the promise of sports analytics. By judging an athlete's performance through good data — as opposed to reputation, image, or outworn clichés — analytics creates the possibility that people can be judged more consistently on merit than often occurs elsewhere in life" (Dizikes, 2021). The extensive use of analytics is no shock for the

younger generation. Analysis of sports data has proliferated as access to analytics capacity has spread in the form of large data sets with decades of player statistics and game outcomes. Proponents of sports analytics are enthusiastic about its use in the present and future. “Embrace data,” said superstar of U.S. women’s hockey Hilary Knight. “It’s here, and it’s the future.” Analytics has diminished some elements of chance in sports regarding the predicted success of up and coming athletes. Andrew Friedman, president of operations for the Los Angeles Dodgers, stated: “Fifteen years ago you saw a lot more bad bets happening a lot more frequently” (Dizikes, 2021). The optimization of sports continues to proliferate, but this calls into question the ethics of implementing such practices.

Golf and GameForge: Innovative Analytics for Recommender Systems

A Technical Report submitted to the Department of Systems Engineering

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Thomas Lee Twomey

Spring, 2022

Technical Project Team Members:

Rose Dennis

Rachel Kreitzer

Jerry Lu

Zachary Kay

Sam Roberts

Steven Wasserman

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

Technical advisor: William T. Scherer, Department of Systems Engineering

Golf and GameForge: Innovative Analytics for Recommender Systems

Rachel Kreitzer, Rose Dennis, Steven Wasserman, Zachary Kay, Jerry Lu, Sam Roberts, Thomas Twomey, William Scherer

Department of Engineering Systems and the Environment, University of Virginia, Charlottesville, United States
rak2cy@virginia.edu, rad7wgg@virginia.edu, stevenwasserman@virginia.edu, zmk2ud@virginia.edu, zl5bf@virginia.edu, scr3ar@virginia.edu, tlt7pg@virginia.edu, wts@virginia.edu

Abstract **–The college sports industry has grown tremendously over the past decade, with NCAA athletic departments recruiting almost half-a-million students to 19,866 teams in 2019 and generating $18.9 billion of revenue the same year. Identifying and selecting the best student-athletes is critical to maintaining the power of these sports programs, aggrandizing the recruitment pipeline and necessitating the demand for novel use of existing technologies. Sports analytics is one response to these growing needs, as its primary use in junior recruitment has presented fruitful for college basketball and football teams across the nation. Golf analytics firm GameForge aims to provide the same insights to college golf coaches, streamlining the recruitment of junior golfers to U.S. universities from around the world. GameForge seeks to develop a two-sided recruiting system that provides insights to junior players and their coaches as well as strengthen its predictive models with the inclusion of new data. A systems-based approach was taken to develop data-driven machine learning models that would provide (a) a proprietary ranking system that compares junior athletes to one another; (b) a relative SWOT analysis that highlights each player's strengths and skill gaps; and (c) a recommender system that suggests potential recruits to college coaches and recommends colleges of best fit to junior players.**

Keywords – sports analytics, student-athlete recruitment, big data modeling, systems integration

I. Introduction

Emerging digital transformation in the sports industry has escalated the role data analytics plays in recruiting and maintaining talented players across a variety of sports. The $620 billion global sports industry is accelerating faster than the entire global gross domestic product, arising from innovative customer experiences that take advantage of consumer technology and broad access to Internet connectivity [1][2]. Professional sports leagues like the National Football League and the National Basketball Association now capture fan engagement through over-the-top (OTT) platforms that offer live streaming, virtual reality experiences, and social media content on personal mobile devices, "leverage[ing] digital media to build direct connections with fans… [and] broaden content reach for sports organizations" [3]. More recently, new companies such as FanDuel and DraftKings have sought to capture market share in the $165-billion American sports betting industry that yielded $44 billion during the pandemic in 2021 [4][5].

The college sports industry is no stranger to this explosive growth – the U.S. Department of Education reported $14.4 billion in revenue for American colleges in 2019, an increase of approximately $750 million every year since 2004 [6]. The National Collegiate Athletic Association (NCAA) is one of the most powerful sports organizations in the country, whose top twenty-five programs are projected to grow in revenue by 116% over the next ten years, a factor more than double the NBA, NFL, NHL, or MLB [7]. And following regulations regarding name, image, and likeness (NIL) recently passed by the NCAA and upheld by the Supreme Court, student athletes have found new opportunities to promote themselves in a burgeoning college sports sponsorship market valued at $100 million, where athletes can earn $1,000 to $10,000 on average annually [8][9].

While private industry and policymakers rush to keep up with the ever-expanding student athlete market, colleges across America are employing data analytics to recruit and retain top talent to NCAA teams. This is further extrapolated between sports of different apparent retail values – U.S. universities spend far more on recruitment in football and basketball compared to other sports because of the demonstrated difference in consumer demand [10]. Junior athletes in other sports, like golf, must rely on specially segmented platforms, like rankings published by the American Junior Golf Associations (AJGA), to demonstrate their value to recruiters.

GameForge, a golf analytics firm, provides a data-driven platform that seeks to ameliorate the junior recruiting process by streamlining information sharing between junior players, college players, and collegiate coaches [11]. Currently, the company offers college students and coaches an online portal that features thorough athlete analyses comprised of relevant descriptive statistics and golfer rankings comparable to different college conferences [12]. However, with the apparent market opportunity entertained by new options for student-athletes and by the sports industry at-large, GameForge seeks to expand their services to better serve the recruitment of junior players. In conducting research in coordination with GameForge, our objective is to develop complex statistical inference and machine learning models that can deliver insight on identifying and recruiting junior golfers as well as provide strategic guidance on the development of a two-sided recruiting system for both junior and collegiate stakeholders. As the magnitude of the college sports industry rises, GameForge can deliver unique golfer tracking capabilities that proffers novel sports analytics techniques and manages junior recruitment practices for its customers.

II. Background

The current approach in collegiate golf recruitment overlooks many golf players that have the potential to improve team performance. Top golfers are easily identified at tournaments and other major golfing events, but mid-level players are rarely considered due to the absence of a tangible platform to demonstrate their strengths. In addition to this,

there is no current way for players to identify teams that are good matches based on metrics beyond rank, such as qualitative factors and personal preferences. This results in both colleges losing out on players that may strengthen their team and players not being able to find a team that will foster their skills and optimize their performance. The absence of a centralized setting that addresses the current recruitment concerns led GameForge to develop a data-driven platform. GameForge currently provides features that allow its users to understand their individual performance and identify training needs. We outlined three specific features to help improve the college golf recruiting experience - a high school player ranking, a method of outlining a player's specific strengths and weaknesses, and a college recommender system for matching junior players and collegiate coaches.

A. Player Rank

Current popular golf associations, such as the American Junior Golf Association (AJGA) and Golfstat, are the standard for ranking players. However, these ranking systems do not allow for direct player comparison across different associations from junior golf to college to the Professional Golfers' Association (PGA) tour [13]. It is a common complaint amongst college coaches that current rankings do not fully capture all talent and potential in the player recruitment pool [14]. Our objective was to develop a proprietary ranking that outperforms the current systems, while allowing coaches to compare an individual player to the current recruitment pool and obtain a projected college rank based on player performance with less bias than current ranking systems [15].

B. Player-Field Performance

An inherent part of comparing athletes is to consider their specific strengths and weaknesses. No two sports players are the same or play their sport the same way. A challenge for many sports analysts is to quantify the strengths and weaknesses of different players to compare them overall. The approach of identifying a golfer's individual skill sets has been brought to golf on a limited scale at the PGA Tour level; however, their statistical measures are not practical for golfers at the high school and college level [16][17]. At the high school and college level, metrics to identify the strengths and skill gaps of a golf player or team do not exist. Coaches that express interest in a specific player often use qualitative decision factors to pinpoint player strengths. This results in golf players being overlooked and players not always committing to a college where their skill set could be optimized. The goal was to provide quantitative metrics that objectively identify how players perform compared to industry levels and other players by using hole variances and means of individual players. Similarly, utilizing proprietary GameForge metrics for driving, irons, short game, and putting gameplay aspects of a player's performance allows the system to identify specific areas to target for improvement. Identifying skills and skill gaps in comparison to the current field allows coaches to analyze specific components of a player's performance. Coaches are then given the opportunity to identify their overall team skill gaps and recruit players that may fill the existing skill gaps.

C. Player Recruitment

Collegiate golf recruiting, like many other university level sports, is a fragmented and inefficient process for both coaches and athletes for several reasons. There is misunderstanding in the requirements to be recruited, poor communication between golf players, recruiters, and coaches, and most importantly, absence of a centralized setting for addressing these issues [18]. Current recruitment for junior golf players consists of creating an online profile, contacting college coaches, competing in tournaments that will gain them recognition, and potentially hiring a private consultant [19]. This creates a confusing, labor-intensive, and sometimes expensive process that can be incredibly overwhelming for high school athletes. In addition to this, it is difficult for coaches and players to identify mutual interest based on player performance and preferences. The objective was to identify various factors that go into selecting a college and generate a list of potential player and college pairs. This will serve to reduce stress and streamline the recruiting process for both players and coaches. Various factors that could impact an individual's choice to commit to a college were explored: student body size, college golf team rank, distance from hometown, geographic regions, social factors and academic factors [20][21][22].

D. Previous Work

GameForge has been working in past years to enhance their analysis and add new features to their platform in order to better serve their users [11][23]. Previous research efforts utilized disparate datasets without clear organization or accessibility, in stark opposition to the now available GameForge database. The GameForge database includes player tournament scorecards for AJGA and PGA tours; rankings from AJGA, Golfstat and WAGR; proprietary, user-inputted GameForge metrics; and collegiate team and player information. With this new resource, the objective was to aid GameForge by generating data-driven insights to provide players and coaches metrics beyond current ranking systems and prestige when committing to a team.

III. Player Rank: Proprietary GameForge Ranking

In chess, the Elo system allows for direct comparison of any two players by their rating [24]. In tennis, the ATP point system provides a numerical system to compare performances within the calendar year [25]. Current golf rankings, however, lack features that allow for head-to-head player comparison while capturing player performance variability due to segmentation of tournaments and rankings. AJGA, for example, only includes tournaments that are invitationals, open tournaments, senior events, all-star series, and preview series [26]. To combat these issues, we developed current rank and projected college rank using GameForge metrics so that current player performance and future potential can be measured more accurately. Both newly-developed, proprietary GameForge ranks outperform the leading industry rankings generated by AJGA after analysis.

A. GameForge Current Rank and Projected College Rank

TABLE 1. Overview of Predictive Models

Player Name	Tournament Outcome	Golfstat	GameForge
Player A	4	5	1
Player B	5	4	2

Player Name	Tournament Outcome	Golfstat	GameForge
Player C	3	1	3
Player D	6	12	4
Player E	10	8	5

*For privacy reasons, player names have been obfuscated

Using GameForge metrics and player scorecards, a stepwise regression model was created to determine significant golf metrics and generate an index-based scoring model for ranking players. The regression was completed using historical GameForge metric data as the independent variables and the latest tournament score as the dependent variable. The model computes factor loadings on the significant metrics and creates a weighted sum that results in an index score for each player. These index scores are then ranked to generate the GameForge current rank, which is organized as a "1224" standard competition ranking (SRC). For head-to-head comparisons, higher-ranked players have greater scores and are estimated to outperform a lower-ranked player. Analysis of the GameForge current rank found it outperformed 20% better than published AJGA rankings and 17% better than Golfstat rankings, as exemplified in TABLE 1. . The same regression methods employed to create current rank were employed to develop the projected college rank. The change was the dependent variable: college ranking. The model for projected college rank accurately predicts the top 25 players with greater than 70% accuracy.

B. Dynamic Rankings

For both rankings, as new tournament data is available, the GameForge metrics are recalculated with the added scorecards, leading to different factor loadings. The dynamic nature of the factor loadings allows the current rank and predicted college rank to better capture variability in performance and predict head-to-head player comparisons more accurately than existing golf ranking systems.

IV. Player-Field Performance: SWOT Analysis

Another important aspect of evaluating players is examining their performance throughout golf rounds to scrutinize their beneficial functional strengths and hindering skill gaps. Collegiate coaches often face challenges in creating well-functioning teams for tournaments arising from a lack of tools that evaluate combinations of golfers in a simple manner. Additionally, traditional golf research does not provide comprehensive feedback to players on their strengths as well as potential areas for improvement [16]. The player-field performance tool provides a succinct overview of each player's course performance through a transfigured SWOT analysis that examines mean score for each par as player strengths and weaknesses as well as unique GameForge metrics as player opportunities and threats.

This tool accomplishes two distinct goals. First, it provides quantitative information for coaches to analyze both their teams and their potential recruits. A coach could analyze their team and see if all their players have a specific strength or weakness; if there are no players who meet a threshold for a current criterion, that could be an important factor they could use when recruiting players for the next year. In addition, it could allow them to shape the lineups for their current team; if the coach knows that a specific type of hole is prevalent or an aspect of the overall golf game is especially important in an upcoming tournament, then they could look at which of their players are strong in those fields when determining the golfers to that tournament. In addition, the tool allows individual players to identify their own strengths and weaknesses to better target areas for training. Since this data is also available to the player to which it pertains, they can see where their game may be lacking and practice specific skills that can help raise their scores.

A. Par Performance

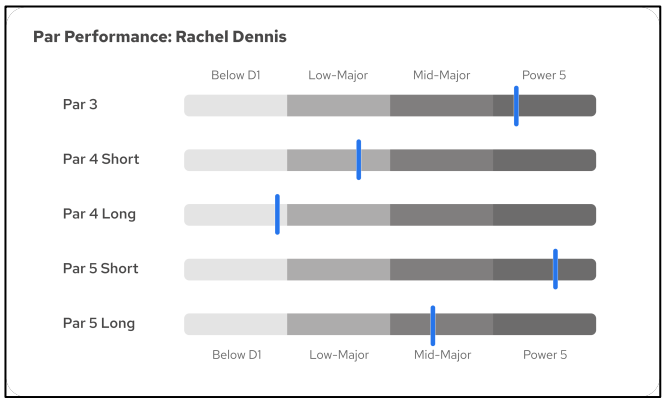

Fig. 1. Analysis of Mean Score Relative to Par for Generic Player, compared to Conference Thresholds

Player performance relative to par is determined using data acquired from high school and collegiate tournament performance, divided by player and subdivided by the hole par associated with the score. This subdivision is required for accurately evaluating a player's consistency in scoring relative to the average number of strokes expected on a given hole. Both junior and college golfers will typically play holes with a par of 3, 4, or 5. Concurrent with previous analyses of the effects of hole yardage on player score relative to par, designations of "short" and "long" for par 4 and 5 are used for holes shorter and longer than the calculated mean yard length, used throughout analysis and shown in Fig. 1 above [11].

After adjusting for player improvement in college, we compare a player's current performance to calculated benchmarks in order to separate them into one of four skill levels for each of the five types of holes. Players are compared to four major categories, including: the Power 5 Conferences schools, which incorporate the most elite conferences of the NCAA; the Mid-Major schools that are considered the "middle-of-the-pack" colleges in Division 1; and the Low-Major schools, which reflect the less competitive Division 1 colleges in the NCAA. Fig. 1 displays a typical analysis of the player's consistency in scoring par for each hole type and includes additional information that allows users to analogize each player to established thresholds of collegiate performance. Similar analyses are performed on the average scores of an entire college team, which evaluates players within teams and

determines frequencies of player types represented on a given team.

B. Players Skills and Gaps

Comprehensive comparison of player performance relative to par is beneficial to recruiting and sustaining competitive golf teams, but it does not aid in directly improving player skills through regimented practice and directed training. GameForge maintains sixteen proprietary metrics that are inputted by users and describe diverse player skills. Metrics are categorized into four areas relevant to different aspects of gameplay (driving, irons, short game, and putting), and aggregate scores are calculated for each. These insights suggest areas for improvement for players as well as competitive thresholds for performance on a skill-by-skill basis. Furthermore, it allows collegiate recruiters to distinguish key attributes from one player to another, supporting colleges in developing rosters of diverse talent.

V. Player Recruitment: Recommender System

Due to the fragmented golf recruitment process, it is important that efforts and resources are directed where there is mutual interest between players and coaches. The recommender system provides players and coaches quantitative confidence when pursuing a potential commitment and gives guidance to both parties during the recruitment process. The system is based on a multifactor model that incorporates various elements a player may consider when selecting a school. In addition, the recommender system integrates junior player strengths and skill gaps from previous analyses to bolster recruitment decision-making, by providing insight into the utility a player and collegiate team can provide each other.

A. Phase I: Individual Predictive Models

Five machine learning models, summarized in TABLE 2. predict various factors that a player considers when selecting a college. The data used to develop these models came from either the GameForge database or was collected from an outside source. The GameForge database data includes player tournament scorecards and AJGA rank. AJGA ranks are composed of junior players who have competed in at least six premier junior golf tournaments in the United States. Other data acquired include hometown size, hometown location, Niche grades, and National Golf Foundation data.

Niche Schools Rankings are a widely recognized college rank system that generates an overall grade for each college based on student survey data. The Niche grades include factors such as academics, athletics, social life, diversity, and safety. Niche grades range from D to A+ with D being the worst and A+ being the best. After testing several binning methods for Model 1, the Niche grades were binned into a high-grade bucket (A+, A, and A-) and a low-grade bucket (B+ and lower). Each athlete's AJGA rank, hometown location, and hometown size were used as independent variables to predict the Niche grade bucket of the college that the player will attend. The best performing method to predict Niche grade was a random forest model. Methods similar to those employed for Model 1 were used to create Models 2 through 5. Each of the models predict values that are indicative of which college a player will select, and all models were fitted using ten-fold cross validation.

TABLE 2. Overview of Predictive Models

Model	Predicting Values	Method	Data
Model 1: Niche Grade	High: A+ to A- Low: B+ and lower	Random Forest	Niche Grade, AJGA rank, hometown location, hometown size
Model 2: Geographic Region	South West Midwest Northeast	Random Forest	Player scorecards, AJGA rank, hometown location, hometown size
Model 3: College Size	Students: < 3000 3000-10K > 10,000	Voting Ensemble: Random Forest, Rule Induction, kNN	Player scorecards, AJGA rank, hometown location,
Model 4: Team Rank	<50 50-100 100-150 >150	Voting Ensemble: Random Forest, Rule Induction, kNN	Player scorecards, AJGA rank, hometown location, hometown size
Model 5: Distance from Hometown	< 250 Miles < 250 Miles	Voting Ensemble: Random Forest, Naive Bayes	AJGA rank, Number of holes played, hometown data

B. Phase II: Multi-Factor Model

A generalized linear model was created using the model outputs from A. Phase I: Individual Predictive Models. Second order interaction terms were significant but did not add predictive power when suggesting player-college pairs. The model outputs scores for all colleges a player can attend and then recommends the schools with the top 15 scores. The list generated by the final multi-factor model accurately captures the college a player attended 80% of the time. The multi-factor model is uses optimizing data that describes where high school students have attended college in the past. To account for player preferences and constraints that were not analyzed, the recommender system would be implemented with the option for a player filter based on the predictive values determined.

VI. Deliverables And Outcomes

A. Proposed GameForge Dashboards

The proposed GameForge Dashboard utilizes the information generated by the machine learning models outlined to aid coaches and junior players. The dashboard would consist of four components: a player profile, college profile, player recommender, and a college recommender. The player profile is an overview of player metrics which includes current rank, college predicted rank, and player-field performance comparisons. The college profile displays the same information as the player profile but metrics of members on a given college team are aggregated. The player recommender suggests junior players to college team coaches based on the multi-factor model as well as player-field performance strengths and skill gaps. The college recommender uses the same information as the player recommender, but conversely suggests colleges to junior players.

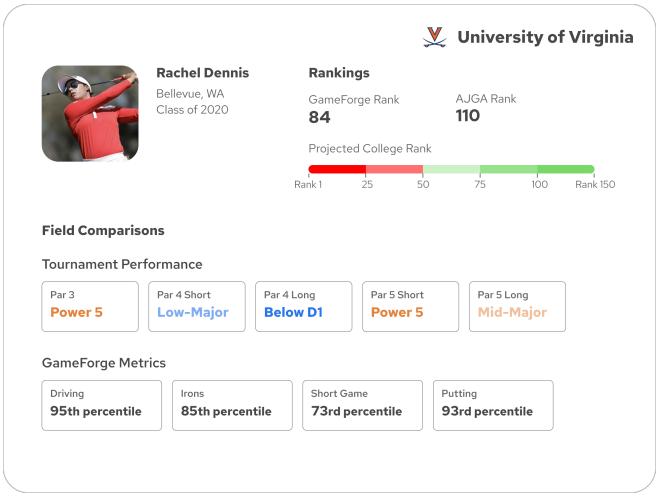

Fig. 2. Player Profile Dashboard

In Fig. *2* above, the tournament performance section indicates the level at which the player performs for par 3, 4 and 5. There are four levels of performance: power 5, mid-major, low-major, and below D1. Rachel Dennis' strengths include that she plays par 3 and par 5 short at a Power 5 level. The "GameForge Metrics" section at the bottom of the dashboard summarizes the player's relative percentile, compared to all junior players, for each of the four categories of GameForge Metrics: driving, irons, short game, putting. Rachel Dennis' driving and putting metrics are above the 90th percentile compared to other players in the field, signifying those categories as her strengths.

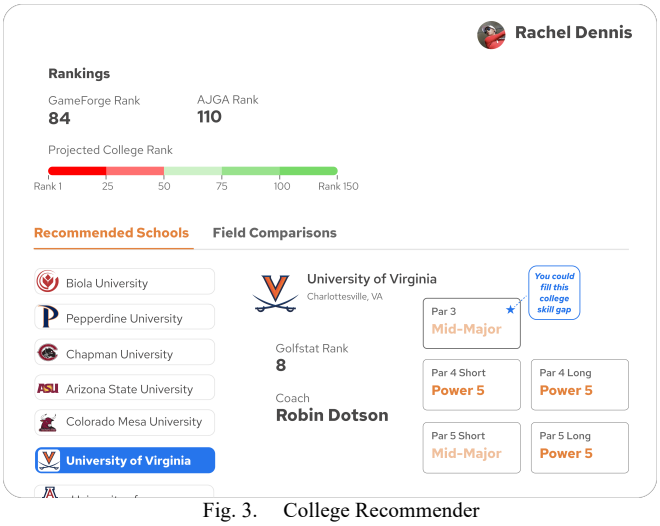

Fig. 3. College Recommender

In Fig. *3* above, the player college recommender displays the top 15 colleges recommended by the multi-factor model to the player. For all of the recommended schools, a player can view the college team's information, strengths, and skill gaps. The sixth best match recommended for Rachel Dennis was the University of Virginia. The blue star icon for par 3 indicates that Rachel, who plays par 3 at the Power 5 level, could fill UVA's par 3 skill gap as the team performs at a mid-major level.

B. Sandbox

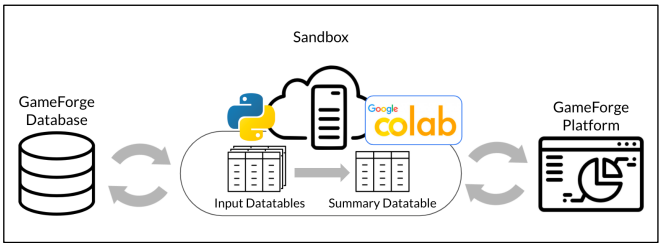

Fig. 4. Sandbox cloud application interacting with GameForge systems

One key distinction in our research and development compared to previous years has been the employ of controlled data management in model research and development. In previous work between the Department of Engineering Systems and the Environment at UVA and GameForge, research relied on disparate datasets to develop and operate statistical models through instance-based execution [11][23]. Since then, GameForge compiled a MySQL database to house related data that could be accessed on an ongoing basis. This inspired the development of Sandbox: a fluid, dynamic environment where the statistical and machine learning models generated as part of this research could be re-run at any point in the future, providing GameForge the opportunity to recalibrate models based on new data and permit ongoing data monitoring without complete instantiation of the models. A simplified Extract-Transform-Load (ETL) pipeline was initialized in a Google Colab file using Python, and models created during the research period were re-created accurately within the environment. Sandbox calls on the GameForge MySQL database, re-runs the models after loading information onto the platform, then pushes data back to the database. This end-to-end product provides the backend information aggregation necessary to compute values given in the proposed GameForge dashboard, and benefits players and coaches alike with streamlined, up-to-date institutional knowledge.

VII. Conclusion

A. Discussion

Through machine learning model creation and product deliverable development, we discovered that junior golfer performance could be modeled to predict eventual collegiate recruitment and college player scoring could be analyzed to increase tournament success. The proprietary GameForge ranking system provides a unique classification of golfers that compares selected junior players to the entire recruiting pool and predicts eventual college performance. The player-field performance analysis identifies player strengths and weaknesses through aggregated mean par scoring as well as opportunities and threats through targeted golf metrics that describe hole performance, both to bolster athlete training by recommending areas for improvement and enhance player recruitment by recommending players with distinct characteristics that can field diverse collegiate lineups. Finally, the player recruitment system combines a variety of descriptive data, including player performance metrics, university ranking factors, and geographic information, to

match players with colleges, aiding both college coaches and junior players in finding the best fit for college teams.

The dynamic interaction between these data-intensive systems provides a wide-ranging, comprehensive view of the field of golf players that permits GameForge users access to key insights on field-wise performance. Conversely, the interwoven use of data allows for a narrow view on an individual basis for close scrutiny of player strengths and skill gaps that can dictate training and recruitment.

B. Limitations and Future Work

When considering the recommender system multi-factor model, incorporating additional factors, such as weather, that a student athlete might consider when selecting a college could increase accuracy in matching players and college. One limitation of the recommender system is the unavailability of personal information about the athletes, such as SAT score or family history, which could provide more insight into school selection. Due to privacy concerns, this data is unattainable, but potentially in the future, athletes using GameForge could opt into providing this kind of information to improve their college recommendations as well as future golf prospects through more historical data. One next step for player-field performance could be to quantify the consistency of each player. Golf is characterized by exceeding amounts of variance from round to round, so player consistency could be an important metric for coaches to consider. One limitation of the player skill gaps methodology is its reliance on user-inputted data. This data is limited to players who are users of the GameForge system and input their own data for each of the proprietary metrics, which results in less data than that found online of all golf players and is subject to self-reporting errors.

ACKNOWLEDGMENT

We would like to extend our sincerest appreciation to GameForge for sponsoring this project, and to Mark Sweeney and Brian Bailie for their continuous support, advice, and engagement over the past eight months.

REFERENCES

[1] Kearney. (2022, April). *The sports market.* Kearney. https://www.kearney.com/communications-media-technology/article?/a/the-sports-market

[2] Hsia, R. (2018, June 28). *The long game – sports analytics and social media insights.* Harnham. https://www.harnham.com/us/post/2018-2/the-long-game-sports-analytics-and-social-media-insights

[3] Giorgio, P., Ohri, L., & Marzin, K. (2018). *A whole new ball game: Navigating digital change in the sports industry.* Deloitte Consulting LLP. https://www2.deloitte.com/us/en/pages/technology-media-and-telecommunications/articles/digital-transformation-and-future-changes-in-sports-industry.html

[4] How much money do americans bet on sports? (2022) *LegalSportsBetting.com*. https://www.legalsportsbetting.com/how-much-money-do-americans-bet-on-sports/

[5] Yakowicz, W. (2021, August 10). *U.S. gambling revenue to break $44 billion record in 2021.* Forbes Media LLC. https://www.forbes.com/sites/willyakowicz/2021/08/10/us-gambling-revenue-to--break-44-billion-record-in-2021/?sh=7c8d1685677b

[6] Office of Postsecondary Education. (2019). *Equity in athletics data analysis* (17th ed.). U.S. Department of Education. https://ope.ed.gov/athletics/Trend/public/#/answer/6/601/trend/-1/-1/-1/-1

[7] Barakat, C. (2017, June 21). The business of athletic recruiting. *Sparks Rowing.* https://sparksrowing.com/blog/the-business-of-athletic-recruiting

[8] Derdenger, T. (2022, February 10). Can you quantify the economic worth of celebrity endorsements? *Carnegie Mellon University Tepper School of Business.* https://www.cmu.edu/tepper/faculty-and-research/research/videos/celebrity-endorsements-tim-derdenger.html

[9] Maestas, A., & Belzer, J. (2020). *How much is NIL worth to student athletes?* AthleticDirectorU. https://athleticdirectoru.com/articles/how-much-is-nil-really-worth-to-student-athletes/

[10] Wittry, A. (2019) *An analysis of college football recruiting costs.* AthleticDirectorU. https://athleticdirectoru.com/articles/an-analysis-of-football-recruiting-costs/

[11] Bassilios, M., Jundanian, A., Barnard, J., Donnelly, V., Kreitzer, R., Adams, S., & Scherer, W. (2021, April 29-30). *Developing a recommendation system for collegiate golf recruiting.* 2021 IEEE Symposium on Systems and Information Engineering Design, Charlottesville, VA, United States. https://doi.org/10.1109/SIEDS52267.2021.9483777

[12] Golf Academy at Heritage Point (2021). Future junior champions. Paul Horton Golf. https://www.paulhortongolf.com/future-junior-champions/

[13] Leivenberg, W. (2013, August 14). *Breaking down the biggest problems with the official world golf rankings.* Bleacher Report, Inc. https://bleacherreport.com/articles/1738273-breaking-down-the-biggest-problems

[14] Johnson, A. (2017, January 7). *A systematic problem: Golf digest's golf course rankings.* The Fried Egg. https://thefriedegg.com/bad-golf-digest-rankings

[15] Broadie, M., & Rendleman, Jr., R. (2013, February 26). Are the official world golf rankings biased? *Journal of Quantitative Analysis in Sports, 9*(2), 127-140. https://mba.tuck.dartmouth.edu/pages/faculty/richard.rendleman/docs/owgr_20130226a.pdf

[16] Broadie, M. (2012). Assessing Golfer Performance on the PGA Tour. *Interfaces*, *42*(2), 146–165. http://www.jstor.org/stable/41472743

[17] Lutes, M. F. (2012). *Power vs. precision: How have the determinants of PGA tour golfers' performance-based earnings evolved since the 1990's?* [Master's thesis, Dalhousie University]. http://hdl.handle.net/10222/15439

[18] Three major mistakes commonly made in todays' college recruiting process (2019, May 31). *ForeCollegeGolf.* https://www.forecollegegolf.com/blog-posts/three-major-mistakes-commonly-made-in-todays-college-recruiting-process

[19] Men's college golf recruiting guide. (2021). *NCSA Sports*. https://www.ncsasports.org/mens-golf

[20] Czekanski, W. A., & Barnhill, C. R. (2015). Recruiting the student-athlete: An examination of the college decision process. *Journal for the Study of Sports and Athletes in Education*, *9*(3), 133-144.

[21] Pauline, J. (2010). Factors influencing college selection by NCAA division I, II, and III lacrosse players. *ICHPER-SD Journal of Research*, *5*(2), 62-69. https://files.eric.ed.gov/fulltext/EJ913334.pdf

[22] Huntrods, K. (2019). *Analysis of factors influencing college choice decisions of mid-major NCAA division I swimmers*. [Master's thesis, South Dakota State University]. https://openprairie.sdstate.edu/etd/3171/

[23] Rohrer, K., Ziller, J., Flores, A., Scherer, W., Kaylor, C., Jimenez, O., & Adams, S. (2020, April 24) *Developing state-based recommendation systems for golf training*. 2020 IEEE Symposium on Systems and Information Engineering Design, Charlottesville, VA, United States. https://doi.org/10.1109/SIEDS49339.2020.9106646

[24] Kovalchik, S. (2020). Extension of the elo rating system to margin of victory. *International Journal of Forecasting*, *36*(4), 1329-1341. https://doi.org/10.1016/j.ijforecast.2020.01.006

[25] ATP rankings. (2022) *Association of Tennis Professionals Tour.* https://www.atptour.com/en/rankings/rankings-faq

[26] Rolex AJGA rankings. (2022) *American Junior Golf Association*. https://www.ajga.org/rankings#:~:text=What%20are%20the%20rankings%3F

Undergraduate STS Research

Predictive Analytics in Sports: Offsetting Human Bias

(sociotechnical research project)

by

Thomas Twomey

May 12, 2022

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.

Thomas Twomey

Technical advisor: William T. Scherer, Department of Systems Engineering

STS advisor: S. Travis Elliott, Department of Engineering and Society

Predictive Analytics in Sports: Offsetting Human Bias

In the U.S., how have social groups responded to the development of sports analytics?

Human bias heavily influences sports. For example, fans develop opinions of athletes based on their own anecdotal experiences from media portrayal to personal interactions. On the team side, coaches choose their starting lineups by how players “look”, meaning their overall athleticism, demeanor, and team play. In the box office, teams play higher paid players to protect their investment. Bias unfairly favors young athletes with more exposure and performance. A young athlete with financial resources will have more recruitment success than a similarly talented player from a poor family. All in all, sports is a business with conflicting agendas amongst participant groups. Certain parties have more influence than others, leading to some biases. The National Football League (NFL) contains bias. “There are black quarterbacks in the NFL; these athletes, however, are still rarely placed in the starting position, the leading spot on a football team” (Viklund, 2009).

Some fans claim that sports analytics are overused. Fans miss the excitement of unpredictability in modern Major League Baseball. “There are a lot of fans that do miss stolen bases,” FanGraphs Managing Editor Meg Rowley said on the virtual panel at the analytics conference. "They don’t care that they’re inefficient" (Golen, 2021). Sports “Superfans” embrace data. “As I watch professional hockey teams hire droves of their own superfans to adapt the sports analytics thinking developed in baseball to a new context it’s worth sounding a note of caution” (Cooper, 2015). Data analysis in sports extends beyond the field. “[Sports organizations] can mine sentiment from social media streams to understand what fans are thinking and can use analytics to engage those fans via social channels” (Ricky, 2019).

There are many examples of bias in sports, and many actors that contribute to this bias. For instance, bias among the rule keepers, or referees, is impactful to the final score of games. Leagues test referees for implicit bias that could affect game outcomes. “It is possible to make predictions of the likelihood of future behavior from bias testing that is significantly better than random guesswork” (Petersen, 2020). Analytics companies market their services to teams. Professional teams look for every advantage. “All professional sports teams are looking for any type of competitive advantage they can obtain as can be seen in the recent sign stealing scandal for my beloved Houston Astros” (Disch, 2020). Are scoreless metrics valued less than measurable athletic performance? Analysis can identify many winning factors but not all. Many successful players do not have characteristics that immediately stand out. For example, quarterback Drew Brees, who recently retired from the NFL in 2021, has hall of fame level accolades despite one major drawback: his size. Typically, height is one of the most desired traits for an NFL level quarterback. Drew Brees was the shortest starting quarterback in the NFL for several years that he was in the league. “If height mattered in the NFL, Drew Brees would be a kicker” and despite this, “He has won a Super Bowl, been named to four pro bowls (2004, 2006, 2008, 2009), and was recently named the cover-boy to the Madden '11 video game“ (Thoms, 2021). Not all success is captured by statistics however, “There are many unsolved important problems in sports for which mathematical analysis can provide a solution” (Winston, 2012).

Predictive analytics can help offset human bias, however the use of analytics is not a perfect system and analytical approaches are not without bias. “Analytics creates the possibility that people can be judged more consistently on merit than often occurs elsewhere in life. But that

promise of fairness only goes so far in a sports world shaped by the same social forces as everything else" (Dizikes, 2021). Analysis is only as impartial as those analyzing the data. Analytics cannot replace human judgment. For example, humans must select which metrics to analyze. "We have hundreds of metrics that could be used, but we don't always have a good understanding of when to use which metric" (Wharton, 2019). Performance metrics are not always straightforward. "It's pretty hard to quantify defense with publicly available data," said Alexandra Mandrycky, director of hockey strategy and research for the Seattle Kraken, the NHL's new expansion team. The metric selection process has the largest potential for human bias.

Confirmation bias is the tendency to interpret new evidence as confirmation of one's existing beliefs. An example of confirmation bias in sports is when a bench player is brought into the game to replace a starting player and booed by the crowd. The next play, that same replacement player makes a costly mistake and the crowd groans that the coach should have kept the starting player in the game. There is no quantitative evidence that the bench player was a worse choice than the starting player, however the crowd's immense displeasure stems from their disagreements with the coach's decision in the first place. Selection bias is the error of selecting individuals who are not representative of the target population. Selection bias occurs in sports when coaches continually pick players based on certain physical characteristics. For instance in baseball, coaches want their pitchers to be tall, so their throws have more velocity, outfielders to be fast, so they can catch more fly balls, and shortstops to be the most athletic, so they can make crucial diving plays when it matters. None of these characteristics alone quantitatively prove one player is more qualified than another who doesn't have one of these stereotypical features.

Framing bias occurs when a decision is made based on the way information is presented. In sports, this can arise when one statistic is valued over another. This is exactly what happened in *Moneyball*. Typically in baseball, slugging percentage, the total number of bases a player records per at bat, is one of the most important statistics for a hitter. When Billy Beane discovered that on base percentage, the percentage a player reaches a base each time they stand up to bat, plus slugging percentage is a much better predictor of a player's skill. Confirmation bias, selection bias, and framing bias negatively influence data analysis (Tanuwidjaja, 2021) and therefore sports analytics.

STS Framework: Social Construction of Technology

How do the different views of stakeholders affect the development of sports analytics?

Now that relevant stakeholders as well as the ethical questions surrounding sports analytics have been established, this practice will be analyzed with a common framework. An important theory to keep in mind in regards to this research problem is the social construction of technology (SCOT). This theory argues that rather than technology shaping the actions of humans, it is actually the actions and attitudes of humans that shape the course of technology. SCOT can be separated into two stages: interpretive flexibility and closure. Interpretive flexibility is concerned with the variation in different social groups' interpretations of the problem, solution, and prioritization of any potential tradeoffs that may need to be made. This can be further split up into three categories: relevant social groups, design flexibility, and problems/conflicts. Relevant social groups are the different segments of the makers and consumers of technology, and can be appropriately segmented based on how they interpret the problem at hand. Design flexibility is the potential for differing construction methods of the

technology based on the attitudes of the relevant social groups. Problems and conflicts occur when groups have different interpretations of the criteria that the technology must fulfill. The second stage of SCOT is closure, which encompasses rhetorical closure and redefinition of the problem. Rhetorical closure refers to the desires of the relevant groups diminishing as they see what they view to be an appropriate solution to the problem. Redefinition of the problem occurs when a technology that has been developed ends up better serving a need different from the original problem at hand. Once closure is achieved, it is not necessarily permanent. The formation of new relevant social groups can spur debate about the efficacy of the current technology and bring the problem back to the stage of interpretive flexibility.

Applying SCOT theory to the research problem of sports analytics helps illuminate what is at stake for different groups in regards to this issue. There are several relevant groups at hand. Leagues provide a framework for teams to compete with the interest of providing a strong entertainment platform for sponsors. Teams provide financial backing to players in hopes of performing well and winning championships. Athletes are incentivized to perform well and help their teams win to promote themselves. Leagues want to maximize involvement by catering to participants' needs. For example, the NCAA catered to players by allowing endorsements (Blinder, 2021). Teams are using predictive analytics to recruit players (Kern, 2020). High school players are doing more than ever to be recruited (Fader, 2016) including participating in tournaments to obtain a junior national rank as well as promoting themselves online to be marketable to coaches. Data engineers believe analytics based recruiting focuses more on athletic performance and aims to ignore erroneous factors, but other groups may not see it this way, which leads to potential problems and conflicts. Design flexibility is demonstrated by the varying

weights of different factors included in recruiting algorithms, which may serve the desires of certain groups over those of others. Closure would occur when the dominant relevant group in this scenario is satisfied with the recruiting algorithms and technologies that are available, but as aforementioned, this is in no way permanent, and as more groups are introduced and the needs of existing relevant groups change, so will the general opinions on analytics in sports recruiting. Transparency of the methods used in the recruiting process would allow participating athletes and fans to hold teams accountable for undesired behaviors of confirmation bias, selection bias, and framing bias.

Consequentialism vs. Deontology: Two Contrasting Philosophies

Should we evaluate the morality of sports analytics based on intention or consequences?

Now that the SCOT framework has been established for evaluating human actions in sports analytics, philosophies for evaluating morality will be examined. Consequentialism and deontology are two different philosophies that can be applied to ethically evaluating the research problem. Deontology focuses on the intention of an action, stating that if an action and the intention behind it is good, then it is moral, regardless of any potential consequences of said action. Consequentialism focuses on the consequences of an action, stating that the basis of the morality of an action is whether it produces a good or bad outcome. Both philosophies can shed light on the the morality of sports analytics

From a deontological perspective, sports analytics are ethical because the intention behind it is to make the sport more competitive by selecting higher quality players and removing the human element of recruiting. This is done by creating an algorithm to quantitatively evaluate

players eliminating potential human bias when a coach provides their opinion. In other words, the evaluation of a player's skill is done by a computer instead of a human with inherent bias. However, when looking at it from a consequentialist perspective, some may argue that sports analytics are not always ethical. Any sort of algorithm is going to be biased based on the human element of the person programming it, and this leaves a big chance that recruiting algorithms could be continuously biased against certain types of players that possess characteristics that the algorithm deems to be less important. Overall, the practice of sports analytics is ethical because it aims to decrease the presence of bias in recruiting and evaluation, however it is not a silver bullet solution because no algorithm is perfect.

Conclusion

As analytics are increasingly prevalent, their merits are hotly contested, with vastly different viewpoints coming from early adopters and lingering skeptics. This is especially true with sports analytics, as sports have a set of very passionate stakeholders. From prospective athletes who have spent their whole lives training, to coaches who feel a connection to each and every player, to fans who tune in every week hoping to see another victory from their favorite team, these groups have very different attitudes when it comes to introducing analytics into the game. Data engineers aim to make their programs as unbiased as possible, but it is impossible to completely remove the human element from any form of analysis. This is not necessarily a bad thing, as there are already many avenues for bias to pop up in sports, from referees' decisions to coaches' gut feelings, but it is not inherently good either. Further evaluation of the morality of analytics in sports can be conducted through the lens of different philosophies such as consequentialism and deontology, which can help paint a fuller picture of the problem at hand.

As the capabilities of analytics increase, and their place in sports grows the ethical responsibilities of those behind these programs grow as well.

References

Blinder, A. (2021). College Athletes May Earn Money From Their Fame, N.C.A.A. Rules. New York Times. https://www.nytimes.com/2021/06/30/sports/ncaabasketball/ncaa-nil-rules.html

Cooper, Ian. (March, 2015). The Dangerous Data Fetishes of Sports Analytics. Wharton Magazine. https://magazine.wharton.upenn.edu/digital/the-dangerous-data-fetishes-of-sports-analytics/

Disch, James G. (March, 2020). The Evolution of the Sports Analytics Program at Rice University; WAR – What is it Good For? TAHPERD Journal.

Dizikes, Peter. (April, 2021). Fair ball! Sports analytics reckons with equity. MIT News Office. https://news.mit.edu/2021/sports-analytics-equity-0414

Fader, M. (2016). Coaches take aim at heartache and hardship of early recruiting. ESPN. https://www.espn.com/espnw/sports/story/_/id/15509558/the-impact-early-recruiting-players-coaches

Golen, Jimmy. (April, 2021). Stats Pioneer Bill James: Don't Blame Us for Boring Baseball. US News. Associated Press. https://www.usnews.com/news/sports/articles/2021-04-09/stats-pioneer-bill-james-dont-blame-us-for-boring-baseball

Kern, T. (2020, Sept 28). How Data is Changing the Future of College Football Recruiting. Market Scale. https://marketscale.com/industries/sports-and-entertainment/how-data-is-changing-the-future-of-college-football-recruiting/

Lewis, Michael. (2003). *Moneyball: The art of winning an unfair game*. New York: W.W. Norton.

Petersen, Thomas. (December, 2020). *Fairness, implicit bias testing and sports refereeing. Journal of the Philosophy of Sport*. Volume 48, 2021 - Issue 1.

Ricky, Abhas. (January, 2019). How Data Analysis In Sports Is Changing The Game. Forbes. https://www.forbes.com/sites/forbestechcouncil/2019/01/31/how-data-analysis-in-sports-is-changing-the-game/?sh=614621f43f7b

Steinberg, Leigh. (August, 2015). Changing the Game: The Rise of Sports Analytics. Forbes. https://www.forbes.com/sites/leighsteinberg/2015/08/18/changing-the-game-the-rise-of-sports-analytics/?sh=1bd1d4b54c1f

Tanuwidjaja, Olivia. (January, 2021). Beware of Biases in Data Analysis. Towards Data Science. https://towardsdatascience.com/beware-of-biases-in-data-analysis-accf0cb9b3a

Thoms, Tanner. (April, 2021). Coming Up Short: Why Height is Overrated in the NFL. Bleacher Report. https://bleacherreport.com/articles/385339-coming-up-short-why-height-is-overrated-in-the-nfl

Viklund, Pat. (February, 2009). *Brains versus Brawn: An Analysis of Stereotyping and Racial Bias in National Football League Broadcasts*. Boston College. https://www.bc.edu/content/dam/files/schools/cas_sites/communication/pdf/thesis09.viklUnd.pdf

Winston, Wayne L. (March, 2012). Mathletics: How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball, and Football. Princeton University Press.

How Can We Overcome the Challenge of Biased and Incomplete Data?. (2019, June 05). Knowledge@Wharton. https://knowledge.wharton.upenn.edu/article/big-data-ai-bias/

Undergraduate Thesis Prospectus

Golf Analysis: Improving College Recruit Ranking Accuracy

(technical research project in Systems Engineering)

Predictive Analytics in Sports: Offsetting Human Bias?

(sociotechnical research project)

by

Thomas Twomey

November 1, 2021

technical project collaborators:

Rose Dennis
Rachel Kreitzer
Jerry Lu
Zachary Kay
Sam Roberts
Steven Wasserman

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.

Thomas Twomey

Technical advisor: William T. Scherer, Department of Systems Engineering

STS advisor: Peter Norton, Department of Engineering and Society

General Research Problem

In the U.S., how have sports organizations used analysis to promote fairness?

Ostensibly, the rules of athletic competition give all competitors an equal chance at success. In reality, most sports are far from a level playing field. Analytics offers quantitative evaluations that may supplement or substitute for qualitative evaluations. “Fairness is part of the promise of sports analytics. By judging an athlete’s performance through good data — as opposed to reputation, image, or outworn clichés — analytics creates the possibility that people can be judged more consistently on merit than often occurs elsewhere in life” (Dizikes, 2021). There is no shock involved in using analytics for the younger generation. Analytics has proliferated as access to analytics capacity has spread. Proponents of sports analytics are enthusiastic about it. “Embrace data,” said superstar of U.S. women’s hockey Hilary Knight. “It’s here, and it’s the future.” Analytics has diminished some of the element of chance in sports. Andrew Friedman, president of operations for the Los Angeles Dodgers, stated: “Fifteen years ago you saw a lot more bad bets happening a lot more frequently” (Dizikes, 2021).

Golf Analysis: Improving College Recruit Ranking Accuracy

How can golf recruiting be improved using predictive modeling and analytics?

College recruiters are pressured to pick athletes that will help their team win. Several rankings exist to predict the performance of future collegiate athletes. No popular rankings use sophisticated models such as those found in *Moneyball* (Lewis, 2003). Our capstone team, Systems Engineering students led by Professor William T. Scherer, is working with a golf recruiting company, GameForge, to create/use predictive models to improve teams’ success by refining evaluation and ranking of potential recruits. The team is tasked with enhancing

GameForge's proprietary online recruiting recommender system designed for coaches, college athletes, and junior athletes. The design and construction of a two-sided college golf recruiting recommender system (for players and coaches) will allow GameForge members to use advanced analytics, built upon significant data sources in GameForge and elsewhere, to provide innovative information. The three main objectives of this research are: Creating a recommender system that can identify junior players for schools (as well as delineate best opportunities for junior players), developing an independent junior player ranking system to predict collegiate success proprietary to GameForge, and enabling team-wide evaluation of players including what-if scenario simulations for addition of new players and identification of key player archetypes. Our findings will utilize linear regression and other data analytics techniques to develop models to predict future player performance.

Predictive Analytics in Sports: Offsetting Human Bias?

In the U.S., how have social groups responded to the development of sports analytics?

Bias heavily influences sports. Fans have anecdotal opinions of athletes. Coaches pick players by how they "look." Teams play higher paid players to protect their investment. Bias unfairly favors young athletes with more exposure and performance. A young athlete with financial resources will have more recruitment success than a similarly talented player from a poor family. Sports is a business that contains bias. The National Football League contains bias. "There are black quarterbacks in the NFL; these athletes, however, are still rarely placed in the starting position, the leading spot on a football team" (Viklund, 2009).

Leagues provide a framework for teams to compete. Teams provide financial backing to players. Athletes help teams perform. Leagues want to maximize participation. For example, the

NCAA catered to players by allowing endorsements (Blinder, 2021). Teams are using predictive analytics to recruit players (Kern, 2020). High school players are doing more than ever to be recruited (Fader, 2016). Analytics based recruiting focuses more on athletic performance and aims to ignore erroneous factors.

Some fans claim that sports analytics is overused. Fans miss the excitement of unpredictability in modern Major League Baseball. “There are a lot of fans that do miss stolen bases,” FanGraphs Managing Editor Meg Rowley said on the virtual panel at the analytics conference. "They don’t care that they’re inefficient" (Golen, 2021). Sports “Superfans” embrace data. “As I watch professional hockey teams hire droves of their own superfans to adapt the sports analytics thinking developed in baseball to a new context it’s worth sounding a note of caution” (Cooper, 2015). Data analysis in sports extends beyond the field. “[Sports organizations] can mine sentiment from social media streams to understand what fans are thinking and can use analytics to engage those fans via social channels” (Ricky, 2019).

How is referee bias mitigated in sports leagues? Leagues test referees for implicit bias that could affect game outcomes. “It is possible to make predictions of the likelihood of future behaviour from bias testing that is significantly better than random guesswork” (Petersen, 2020). Analytics companies market their services to teams. Professional teams look for every advantage. “All professional sports teams are looking for any type of competitive advantage they can obtain as can be seen in the recent sign stealing scandal for my beloved Houston Astros” (Disch, 2020). Are scoreless metrics valued less than measurable athletic performance? Analysis can identify many winning factors but not all. “There are many unsolved important problems in sports for which mathematical analysis can provide a solution” (Winston, 2012).

Predictive analytics can help offset human bias. However, those conducting analysis inherently have bias. “Analytics creates the possibility that people can be judged more consistently on merit than often occurs elsewhere in life. But that promise of fairness only goes so far in a sports world shaped by the same social forces as everything else” (Dizikes, 2021). Analysis is only as impartial as those analyzing the data. Analytics cannot replace human judgment. For example, humans must select which metrics to analyze. “We have hundreds of metrics that could be used, but we don’t always have a good understanding of when to use which metric” (Wharton, 2019). Performance metrics are not always straightforward. “It’s pretty hard to quantify defense with publicly available data,” said Alexandra Mandrycky, director of hockey strategy and research for the Seattle Kraken, the NHL’s new expansion team. The metric selection process has the largest potential for human bias. Confirmation bias, selection bias, and framing bias negatively influence data analysis (Tanuwidjaja, 2021).

References

Blinder, A. (2021). College Athletes May Earn Money From Their Fame, N.C.A.A. Rules. New York Times. https://www.nytimes.com/2021/06/30/sports/ncaabasketball/ncaa-nil-rules.html

Cooper, Ian. (March, 2015). The Dangerous Data Fetishes of Sports Analytics. Wharton Magazine. https://magazine.wharton.upenn.edu/digital/the-dangerous-data-fetishes-of-sports-analytics/

Disch, James G. (March, 2020). The Evolution of the Sports Analytics Program at Rice University; WAR – What is it Good For? TAHPERD Journal.

Dizikes, Peter. (April, 2021). Fair ball! Sports analytics reckons with equity. MIT News Office. https://news.mit.edu/2021/sports-analytics-equity-0414

Fader, M. (2016). Coaches take aim at heartache and hardship of early recruiting. ESPN. https://www.espn.com/espnw/sports/story/_/id/15509558/the-impact-early-recruiting-players-coaches

Golen, Jimmy. (April, 2021). Stats Pioneer Bill James: Don't Blame Us for Boring Baseball. US News. Associated Press. https://www.usnews.com/news/sports/articles/2021-04-09/stats-pioneer-bill-james-dont-blame-us-for-boring-baseball

Kern, T. (2020, Sept 28). How Data is Changing the Future of College Football Recruiting. Market Scale. https://marketscale.com/industries/sports-and-entertainment/how-data-is-changing-the-future-of-college-football-recruiting/

Lewis, M. (2003). *Moneyball: The art of winning an unfair game*. New York: W.W. Norton.

Petersen, Thomas. (December, 2020). *Fairness, implicit bias testing and sports refereeing. Journal of the Philosophy of Sport*. Volume 48, 2021 - Issue 1.

Ricky, Abhas. (January, 2019). How Data Analysis In Sports Is Changing The Game. Forbes. https://www.forbes.com/sites/forbestechcouncil/2019/01/31/how-data-analysis-in-sports-is-changing-the-game/?sh=614621f43f7b

Tanuwidjaja, Olivia. (January, 2021). Beware of Biases in Data Analysis. Towards Data Science. https://towardsdatascience.com/beware-of-biases-in-data-analysis-accf0cb9b3a

Viklund, Pat. (February, 2009). *Brains versus Brawn: An Analysis of Stereotyping and Racial Bias in National Football League Broadcasts*. Boston College. https://www.bc.edu/content/dam/files/schools/cas_sites/communication/pdf/thesis09.viklUnd.pdf

Winston, Wayne L. (March, 2012). Mathletics: How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball, and Football. Princeton University Press.

How Can We Overcome the Challenge of Biased and Incomplete Data?. (2019, June 05). Knowledge@Wharton. https://knowledge.wharton.upenn.edu/article/big-data-ai-bias/