E-Sports Analytics: A Primer and Resource for Student Research Projects and Lesson Plans

Joseph Rubleske
Northern Kentucky University

Travis Fletcher
Northern Kentucky University

Brett Westerfeld
Northern Kentucky University

Abstract

Electronic sports (e-sports) can be defined as digital games played competitively for an audience (Hodge et al., 2017). With a global consumer base of roughly 450 million people and projected 2019 revenues of US$1.1 billion, the e-sports industry continues to grow (Pannekeet, 2019). Behind this growth is a thriving ecosystem which includes e-sports game publishers, e-sports players and coaches, e-sports content producers, e-sports consumers, and more. As this ecosystem flourishes, so too flourishes an e-sports labor market, a portion of which comprises e-sports analytics jobs.

Demand for e-sports analytics is expected to increase significantly in the coming years in the same way that demand for sports analytics increased after the 2004 publication of Michael Lewis's Moneyball (Davenport, 2014). Given current and anticipated demand for e-sports analytics, some universities might do well to prepare students for e-sports analytics careers. To this end, this article serves as a primer and resource for student research projects and lesson plans around e-sports performance analytics.

This objective is served through four main sections. In the first section, the intelligence needs that fuel the growth of e-sports performance analytics are described. The second and third sections provide practical and detailed information about e-sports data collection. The fourth section draws from the authors' analysis of postings for e-sports analytics jobs to discuss two broad areas of e-sports analytics work: attribute creation for visualizations and dashboards; and machine learning modeling. The article concludes with a brief discussion of popular programming languages and software applications for e-sports analytics.

Keywords: analytics education, student research, lesson plans, e-sports, e-sports analytics

INTRODUCTION

In August 2018, the University of Akron announced plans to phase out 80 degree programs that suffer from low enrollment (Pettit, 2018). In turn, roughly US$6 million became available for reallocation, much of which was earmarked for facilities and programs associated with electronic sports (e-sports), which can be defined as digital games played competitively for an audience (Hodge et al., 2017). To this end, Akron has since debuted three e-sports facilities: one for its varsity e-sports teams; one for its e-sports club teams; and one for its students who play e-sports recreationally (Mitrevski, 2018). Matches played by the varsity teams are displayed in real time on large screens at the flagship facility (Akron Esports Arena), which is located within Akron's football stadium. At the time of this writing, one major sponsor is on board and other partnerships are in the works. Each member of a varsity e-sports team receives a minimum of US$1,000 in scholarship funds, and sometimes more (Bamforth, 2018).

The University of Akron is far from alone in terms of investing substantially in e-sports programs and facilities. The National Association of Collegiate Esports (NACE) includes 144 members as of May 2019 (https://nacesports.org/school-directory). Formed in 2016, NACE comprises staff, a Board of Directors, and a Competition Council; it also claims 11 partners and sponsors and serves as a clearinghouse for job postings. (Job openings include e-sports program directors, coordinators, and head coaches.) NACE also currently administers competitive leagues around four e-sports games: Hearthstone; Overwatch; Paladins; and Smite.

As universities dedicate more resources to e-sports, the global e-sports industry continues to expand. According to recent reports on the state of the e-sports industry, the audience for e-sports content currently includes more than 450 million people globally, up from 395 million in 2018 (Pannekeet, 2019). During March 2019, the five most watched e-sports games on Twitch – namely, League of Legends, Fortnite, DOTA 2, Counter-Strike: Global Offensive (CS:GO), and Overwatch – together claimed roughly 400,000 viewers during any given hour, on average (TwitchTracker, 2019). And according to one report, e-sports will claim a larger audience in the US than every professional sports league except the National Football League (NFL) by 2021 (Iqbal, 2019).

One source projects that global revenues from e-sports will total US$1.1 billion in 2019, up 27 percent from 2018 (Newzoo, 2018). The largest stream of revenues comes, for now at least, from sponsors (US$457 million), media rights (US$251 million), advertising (US$189 million), and tickets and merchandise (US$104 million) (Pannekeet, 2019). In January 2018, Twitch agreed to pay Activision US$90 million over two years for rights to broadcast a majority of Overwatch League content in North America (Baccellieri, 2018). Many other large e-sports broadcasting deals have been made (Newzoo, 2018). Finally, the prize pools for major e-sports tournaments are getting larger; in 2018, the prize pool for the DOTA 2 International – held at the 20,000-seat Rogers Arena in Vancouver, BC – was US$25.5 million (Goldman Sachs Investment Research, 2019). To put this in some perspective, the prize pool for the 2018 Daytona 500 was US$19 million.

E-Sports Games and the E-Sports Ecosystem

For a video game to be recognized as an e-sports game, there must be organized competitions around it, either at an amateur or professional level. Professional e-sports players and teams compete in many types of e-sports games, including:

- Multi-player online battle arena (MOBA) games (e.g., DOTA 2, League of Legends, Heroes of the Storm) and real-time strategy games (e.g., StarCraft II);
- First-person shooter (FPS) games (e.g., Call of Duty, Counter-Strike: Global Offensive, Overwatch);
- Battle royale games (e.g., Apex Legends, Fortnite, PlayerUnknown's Battleground);
- Fighting games (e.g., Street Fighter, Super Smash Bros., Tekken);
- Athletic (aka sports) games (e.g., FIFA, Madden NFL); and
- Digital card games (e.g., Hearthstone).

Professional e-sports team operate in ways that are very similar to how traditional sports teams operate. For example, professional e-sports teams comprise coaches and staff, their players are paid a base salary, and they train regularly and prepare to compete against opponents. Most professional e-sports teams travel to tournaments and matches, and some even have nutritionists and physical therapists on staff (Goldman Sachs Investment Research, 2019). Professional e-sports team also compete with each other to attract and retain the best e-sports players.

Around these professional e-sports games and competitors, a thriving ecosystem has emerged which includes myriad participants: game publishers (aka game developers); owners and coaches of e-sports teams; administrators of e-sports leagues and tournaments; designers and operators of e-sports facilities; e-sports content distributors (e.g., Twitch, YouTube Gaming, Caffeine); e-sports content producers (e.g., streamers, online news producers and aggregators); providers of e-sports gambling services; administrators of fantasy e-sports sites; e-sports consulting firms and recruiters; and, perhaps most importantly, consumers of e-sports content and services. In addition, and much like the traditional sports ecosystem, the e-sports ecosystem depends increasingly on providers of e-sports research and analytics services.

The E-Sports Labor Market and the Need for E-Sports Analysts

As the e-sports ecosystem grows, so too grows the e-sports labor market. Starting in 2018, Syracuse University began working with Twitch to design and offer a recurring course aimed at surveying job and career opportunities in the e-sports industry (Carp, 2018). Further, an analysis of job postings from major e-sports game publishers, e-sports teams, and e-sports consulting firms – conducted by the authors for this study, and described in the Data Analytics section – revealed hundreds of job openings in areas such as software engineering, IT infrastructure management, animation and graphic design, business administration (e.g., human resources, accounting), business operations (e.g., event management, customer service), and business development and marketing.

The same analysis revealed many openings for jobs around e-sports research and analytics. According to several tech-industry articles, the demand for such jobs is expected to increase significantly in the coming months and years (e.g., Bozorgzadeh, 2017; Niederhoffer, 2018; Van Allen, 2018), much in the same way that demand for sports analysts increased after the publication in 2004 of Michael Lewis's Moneyball: The Art of Winning an Unfair Game (Davenport, 2014).

As detailed in the next section, the growing demand for e-sports intelligence will fuel this need for e-sports analysts (Hodge et al., 2017). For example, the analysis of e-sports match and tournament data has already changed the way that competitive e-sports games are played and understood (Van Allen, 2018). Since 2015, if not earlier, some of the best DOTA 2 teams have employed data analysts to help gain a competitive advantage (Summerville et al., 2016).

Ultimately, e-sports teams and players depend on analyses that assess their performance (and thus inform their efforts to improve) and on analyses that assess opponents' strengths and weaknesses. Moreover, e-sports game publishers require analyses that help them improve gameplay by revealing patterns of behavior by users, while providers of e-sports gambling services rely on analyses that predict who will win an e-sports match and by how much.

Given this demand for e-sports intelligence, some universities might do well to offer lesson plans and even elective courses that prepare undergraduate and graduate students for jobs and careers in e-sports analytics. At the same time, such lesson plans and courses may also be of interest to students who simply want to learn data analytics concepts and methods as applied to a timely and increasingly important domain.

Accordingly, this article serves as a primer and resource for lesson plan development and student research projects around e-sports analytics. In the first of four main sections, recurrent intelligence needs around e-sports performance analytics are identified and discussed. In the second section, the three most common types of data sources for major e-sports games are described. The third section discusses three variables (i.e., game patch, skill level, game mode) that must be taken into account (or controlled for) during data collection. The fourth and final section (Data Analytics) comprises a sub-section on attribute creation for visualizations and dashboards and another sub-section on machine learning modeling. The article concludes with a discussion of popular programming languages and software applications for e-sports analytics.

INTELLIGENCE NEEDS AROUND E-SPORTS PERFORMANCE ANALYTICS

Several stakeholder groups have a need for intelligence yielded by e-sports analytics. For example, e-sports game publishers want analytics to help identify patterns in gameplay in order to make changes to game maps, rules, characters, and items (Bozorgzadeh, 2017). They also want analytics to help detect cheating and inform the development of anti-cheating functionality (Lam, 2017). Many e-sports content distributors are investing heavily in analytics aimed at gauging and increasing consumer engagement with online media channels and streams (Newzoo, 2018), while many e-sports content producers are already using analytics to automatically identify e-sports match highlights and produce commentary during and after e-sports matches (Drachen, 2017).

However, the remainder of this article focuses on a particular and prevalent e-sports intelligence need, namely, the need for intelligence around e-sports performance. By "e-sports performance analytics" this article refers to analytics designed to help assess and possibly improve the performance of e-sports teams and players. For example, most if not all professional e-sports coaches want to be able to assess the performance of their team, and of each player on their team, in order to inform improvement efforts. Moreover, these coaches also want to be able to assess the performance of opposing teams and players in order to prepare to defeat them. In this regard, e-sports is much like traditional sports.

Professional e-sports players and coaches are not the only groups who have a need for e-sports performance analytics, though. Specifically, providers of e-sports gambling services, e-sports game publishers, e-sports content producers, and e-sports consumers all have a need for such analytics. For example, e-sports content producers such as writers, streamers, and match commentators draw from e-sports performance analyses to produce more insightful content. At the same time, e-sports fans benefit from e-sports performance analyses as enthusiasts, as gamblers, and as "owners" of e-sports fantasy teams. With these stakeholders in mind, intelligence needs around e-sports performance are organized here around four units of analysis: team; player; character; and item. Each of these units of analysis is described here in turn, together with its associated intelligence needs.

Team Performance and Strategy

A majority of professional e-sports games comprise matches played by opposing teams (e.g., CS:GO, DOTA 2, Heroes of the Storm, League of Legends, Overwatch). Professional e-sports matches played by opposing individuals (e.g., Hearthstone, StarCraft II) are less common. Accordingly, the need to better understand performance and strategy at the team level is critical. From a coach's perspective, team-based performance analyses can (1) aid preparations to compete against an opposing team and/or (2) improve the coach's understanding of his or her own team.

As with traditional sports, analyses which aid preparations to compete against an opposing team yield a scouting report. For most professional e-sports games, team-based scouting reports may address three broad intelligence needs:

1. The need to assess an opposing team's performance. As the head analyst for the Overwatch League's Toronto Defiant remarked, "My job is to identify opponents' weaknesses and exploit them" (Niederhoffer, 2018). What does the opposing team do well? What does it not do well? How can its strengths be countered, and how can its weaknesses be exploited?
2. The need to discover an opposing team's match strategies.[1] Which strategy does an opposing team typically assume at the outset of a match and during a match as conditions change? For example, does the team tend to be more aggressive (offensive) or more passive (defensive)? Does an opposing Overwatch team employ a disengagement strategy? How and when does it use its players' "ults" (i.e., powerful abilities or "ultimates" that take time to charge before use)? Which combination of roles (damage, tank, support) does it employ? Can a strategy for a team's movement inside a given map be discerned? If such strategies can be anticipated, they can be countered.
3. The need to discover an opposing team's character-selection or drafting strategy (where applicable). In e-sports games where players assume the form of a given character – that is, an avatar-like "hero" or "champion" with certain abilities and limitations – professional teams think very carefully about which characters to select (e.g., Kinkade et al., 2015). Specifically, their selections demand that they consider which characters their players are best at using, which characters best complement each other (i.e., promote team synergy), and which characters are best at countering the abilities of characters selected by the opposing team. Accordingly, it behooves teams to try to predict which characters the opposing team will likely select.

   With regard to professional DOTA 2 and League of Legends matches, character selection is more strategically complex because opposing teams engage in a drafting process in which each team "bans" or "picks" a character in each round. (When a character is banned, neither team can select it.) In DOTA 2, for example, teams take turns banning heroes in the first four rounds, then take turns picking heroes in the next four rounds, and so on, until the end of the

[1] The authors' search for germane peer-reviewed articles yielded no article aimed at advancing an e-sports strategy discovery or prediction model. Thus, while coaches and players undoubtedly have a need for such insights, there is currently little if any public scientific knowledge on this front.

draft. Thus, the coaches for DOTA 2 teams and League of Legends teams (which engage in a similar drafting process) have a strong need for intelligence around the opposing team's drafting strategy.

Professional e-sports coaches also have a need for intelligence around the performance and strategies of their own team, as such reports can inform their strategy formation and guide performance improvement efforts. To generate these reports, analysts must address the same types of intelligence needs that are addressed by scouting reports:

1. The need to assess the performance of one's own team.
2. The need to discover the match strategies of one's own team. A coach may assume that she knows the strategies her team employs, but an analysis of match data may reveal that the team actually employs different strategies, or deviates from planned strategies, particularly as unexpected conditions emerge.
3. The need to discover the character-selection or drafting strategy of one's own team (where applicable). Similar to the second intelligence need, an analysis of actual character selection or draft data may confirm that a particular strategy is employed, but it may also reveal that the team does not select or draft characters in an assumed way.

Finally, learning more about an e-sports team's performance also is important as a means of predicting the winner of an e-sports match (Schubert et al., 2016). Indeed, win prediction "has formed the focal point of e-sports analytics," at least to date (Hodge et al., 2017, p. 1). While coaches may have some interest in win prediction, the stakeholders with the most interest in win prediction are providers of e-sports gambling services, e-sports content producers (e.g., commentators, writers, streamers), e-sports fans (as enthusiasts and as gamblers), and e-sports data analysts who want to evaluate the efficacy of machine-learning algorithms as applied to a fairly straightforward problem operationalized with a binary target attribute (i.e., team victory or loss).

Machine-learning models built to predict team victory (or defeat) can be classified according to the unit of analysis on which its training and test datasets are based. Win-prediction models built using only team historical data are the most common (e.g., Semenov et al., 2017; Kinkade et al., 2015; Ong et al., 2015; Yang et al., 2014). These datasets comprise attributes representing a team's performance over some period of time (e.g., earned gold per match, kills per match), such as a season, patch, set of tournaments, etc. A second class of win-prediction models draws not from team historical data but rather from in-match data (i.e., data generated during runtime) and/or pre-match data (i.e., character draft or selection data) (see, e.g., Schubert et al., 2016; Gao et al., 2013). Finally, a third class of win-prediction models draws from both team historical data and from pre-match or in-match data (Hodge et al., 2017; Yang et al., 2016). The latter models may be the least common of the three types.

Player Performance

The need to better understand performance at the player level also is critical, whether the player in question is a member of a professional e-sports team or a competitor in an individual-oriented e-sport (e.g., Hearthstone, StarCraft II). From a coach's (or player-coach's) perspective, player-based performance analyses can (1) aid preparations to compete against an opposing player, (2) improve the coach's understanding of her own players (or the player-coach's understanding of herself), and/or (3) assess a prospective player. Analyses which aid preparations to compete against an opposing player yield a player-centric scouting report that is similar to a team-based scouting report in terms of the intelligence needs it addresses. Specifically, what does the opposing player do well? What does she not do well? How can the player's strengths be countered, and how can her weaknesses be exploited?

It is quite important to consider, though – at least for e-sports games where players assume certain roles and play as certain characters –that a player's strengths and weaknesses may be closely tied to their roles and/or characters. For example, a certain professional DOTA 2 player may be more skilled at assuming a "carry" role (rather than, say, a "support" role); moreover, he may be more adept at playing certain carry heroes (e.g., Spectre) and not others (e.g., Ursa). Thus, a more advanced analysis of the e-sports player might investigate the player-role or player-character as the unit of analysis. In Yang et al. (2014), for instance, certain attributes associated with a DOTA 2 player-role (e.g., number of times player deals damage first during combat for the "initiator" role, number of opposing heroes killed for the "ganker" role) were found to be more effective predictors for team victory.

Analyses aimed at better understanding one's own players (or oneself) can guide individual improvement efforts. There is a dimension to the assessment of a coach's own players that is not present in the assessment of opposing players, though, and that dimension centers on the ongoing need for player evaluation. How well does the player contribute to the team's success? Does the player merit a feature role on the team, or should she be an alternate? If the player is a professional, is he worth his salary? Along these lines, the need for analytics aimed at individual assessment and skill improvement has led to the growth of e-sports coaching and training software. For example, several companies (e.g., DOJO Madness) offer virtual coaching and training software that provides personalized and actionable analysis for certain e-sports games.

Finally, player-based performance analyses also can be used to assess prospective players. In traditional sports, analysts employed by a professional team analyze nearly every player who could potentially merit recruitment, including collegiate players and players in other leagues (e.g., minor leagues, European or Asian leagues, etc.). Competition for potentially talented recruits can be intense and constant, and can be a key factor to a team's long-term viability. While some e-sports team do this to some extent (Van Allen, 2018), it appears that prospect assessment is not done to the same extent in e-sports, at least not yet. As e-sports league and team revenues grow, though, prospect assessment will likely become more commonplace and sophisticated.

Character and Item Performance

With regard to the use of analytics to assess and improve performance at the team and player levels, e-sports and traditional sports are quite similar. One important way in which e-sports differs from traditional sports, however, is in the need for e-sports stakeholders, and especially players and coaches, to assess and better understand the "characters" that players assume as well as the "items" that figure centrally in many e-sports games.

As mentioned above in the section on intelligence needs, a character is an avatar-like "hero" (e.g., DOTA 2, Heroes of the Storm, Overwatch) or "champion" (League of Legends) with certain abilities and limitations. In Overwatch, for example, there are 30 heroes to choose from as of this writing. One of these heroes is named Moira, who ostensibly is used in a "support" role rather than in a "tank" or "damage" role. Moira can heal and protect teammates while also dealing damage to opposing heroes; she can also teleport within short distances. Moira's weaknesses include taking a lot of damage from large bursts and not being as powerful, or perhaps as versatile, of a healer as some other support heroes.

Understanding not only a character's abilities and limitations, but more importantly their effectiveness over thousands of competitive matches, where they are used by hundreds of professional players, is crucial to character selection (or draft) decisions and to strategies around how to make the best use of them. While relatively few empirical studies have focused on the character as the unit of analysis, two studies serve as exceptions: first, Summerville at al. (2016) built a neural network-based model to predict hero draft decisions in order to help coaches improve their drafting strategy; and second, Liang (2017) developed a Champion Viability Score (CVS) which builds on the widely used pick-ban rate (i.e., the percentage of matches in which a given champion is either picked or banned) by increasing the weight given to more recent observations. The value of Liang's (2017) CVS lies in its recognition that when there is a competitive advantage to using a certain new or revised character, that advantage fades rapidly as other players and coaches become aware of it.

In some e-sports games, but most notably DOTA 2 and League of Legends, the items purchased by a player during a match, and used at the right time and place, can make the difference between victory and defeat. For League of Legends, there are currently 281 items and four major categories for these items: defense; attack; magic; and miscellaneous. Each item can be grouped into one or more of these categories; the Abyssal Mask, for example, is both a defense item and a magic item. Every item also has a specified cost and one or more unique abilities. Spectre's Cowl, for example, costs 1,200 gold and grants 150 percent base health regeneration for up to 10 seconds after taking damage from an opposing champion.

Very few studies have used the in-game item as a unit of analysis, though there is clearly a need for such research given the importance of item use. One exception is a study in which the aim was to predict which item(s) a player will buy within the next five minutes during a DOTA 2 match (Looi et al., forthcoming). To this end, the authors developed two models – one based on logistic regression, the other based on association rules – which predicted item purchases by highly skilled DOTA 2 players with accuracies ranging from 67 percent to 87 percent, depending on extant conditions.

THREE COMMON DATA SOURCE TYPES FOR E-SPORTS GAMES

Before data can be analyzed, it must be collected. Accordingly, an analyst based in any domain, including e-sports, must be capable of collecting data in a variety of ways from multiple sources. The three most common types of sources from which e-sports data may be collected are application programming interfaces (API), third-party APIs and statistics providers, and replay-file repositories. Each is discussed here in turn with examples.

Application Programming Interfaces

For some major e-sports games, the game's publisher provides an application programming interface (API) through which data generated from gameplay may be downloaded by anyone who has registered for an account and who in turn has received an API "key" which allows the API provider to associate the account holder with her download requests (queries).

The API provider may need to do this in order to impose limits on the number of queries made during a certain time period.

In effect, an API is a uniform resource locator (URL) through which the desired data can be accessed. Included in the URL (query) – which functions as a call or request to a server – are request conditions (e.g., authentication) and one or more parameters (e.g., object identifiers such as match ID, start and end dates) which serve as filters. The results yielded by the query are typically encoded as JSON-based data objects. In most cases, the API provider offers documentation (e.g., data descriptions, example queries) which helps analysts prepare these queries.

Unfortunately, not all publishers of major e-sports games provide a public API, as doing so can consume a great deal of developers' time and effort while yielding only indirect monetary benefits. No public API is offered for Call of Duty (Activision), Fortnite (Epic Games), Hearthstone (Blizzard) Heroes of the Storm (Blizzard), or Overwatch (Blizzard), at the time of this writing at least. However, a public API is available for Counter-Strike: Global Offensive (Valve),[2] DOTA 2 (Valve),[3] League of Legends (Riot Games),[4] PlayerUnknown's Battleground (PUBG),[5] and StarCraft II (Blizzard).[6]

As an example, the following basic query, executed against the OpenDotaAPI,[7] returns attributes associated with each DOTA 2 hero: https://api.opendota.com/api/heroes. An excerpted screenshot, shown in Figure 1 in the appendix, shows the JSON-based output as it appears in a web browser.

It is important to note, though, that most performance-based analyses require multiple API queries as well as data-processing scripts which tie observations together. Thus, the learning curve for API-based data collection can be steep. Consider, for example, a case where one wants to analyze the players who competed in a certain professional DOTA 2 tournament (e.g., the StarLadder ImbaTV DOTA 2 Minor #2). In order to collect attributes related to each player, one must first identify all the matches (by match ID) that constitute this tournament. This query is made difficult, though, because there is no resource that identifies the values for these match IDs. As a result, the analyst must write a script which identifies the most recent match and then iterates through pages of query results until all pertinent match IDs are found. (One of the authors of this article uses the Perl programming language to help collect and process query results.) For the next step, the analyst can write a separate API query for each pertinent match ID – which is a viable approach given the limited number of matches constituting a given tournament or season – and then another data-processing script that extracts the attributes for each individual player who competed in that match.

There may be some straightforward cases where an analyst is able to write a single API-based query which returns the data she needs in the desired form and structure – that is, where the unit of analysis serves as the object in the JSON-based results, and all desired attributes are contained in that object. Where this is the case, the analyst may need only to use basic software (e.g., MS Excel) to convert these results into a CSV file. In most cases, though, collecting e-

[2] For the CS:GO API, see https://developer.valvesoftware.com/wiki/Steam_Web_API
[3] For the DOTA 2 API, see https://developer.valvesoftware.com/wiki/Steam_Web_API
[4] For the League of Legends API, see https://developer.riotgames.com
[5] For the PUBG API, see https://developer.pubg.com
[6] For the StarCraft II API, see https://develop.battle.net/documentation/api-reference/starcraft-2-game-data-api
[7] The public OpenDota API (see https://docs.opendota.com/) is not administered by Valve, but rather by a third party. Nevertheless, it is widely seen as easier to use, and has more extensive documentation, than Valve's API.

sports performance data via API requires some level of expertise gained through reading API documentation and then executing and re-executing queries using a trial-and-error approach.

Third-Party APIs and Statistics Providers

Some APIs that support the collection of e-sports performance data are provided by third parties. A subset of these APIs draw from a public API provided by a game's publisher, while another subset obtains gameplay data through more creative means. Examples of the former include the aforementioned OpenDota API and the League of Legends Data Solution (LDS).[8] Examples of the latter include the Call of Duty Tracker Network API,[9] the Fortnite Tracker Network API,[10] fortniteapi.com,[11] overwatch-api-server,[12] the Hearthstone API,[13] and HotsAPI for Heroes of the Storm.[14] These third-party APIs obtain their data in a variety of ways, including searching local game logs for non-public API endpoints, using packet sniffers to locate replay-file repositories and non-public API endpoints, scraping data from match videos streamed to Twitch and YouTube, and requesting replay files from users. Some third-party API providers may even be granted access to the game publisher's API endpoints through non-publicized agreements with the game publisher.[15]

Other third-party sites offer performance-based statistics but not APIs. These sites may obtain their data from the publisher's API (where available), from a third-party API, and/or from replay files. (Replay files are described in brief below.) They do not offer as many observations and attributes as APIs, though; rather, they typically provide a number of preset tables or "views" around certain units of analysis (e.g., team, player, character). For example, nearly every stats-based site provides a "leaderboard" or "rankings" page that shows, from top to bottom, the game's best teams and/or players according to some numeric measure (e.g., total points earned). Most of these sites also provide basic statistics for these teams and/or players and, where applicable, for characters (i.e., heroes or champions). As an example, Figure 2 (see appendix) presents an excerpt of a view containing statistics for the 10 teams that competed in the North American League of Legends Championship Series (NA LCS) 2019 Spring Regular Season. Examples of these third-party stats-based sites include:

- CS:GO Stats (https://csgo-stats.com) for Counter-Strike: Global Offensive;
- DatDota (www.datdota.com) and DotaBuff (www.dotabuff.com) for DOTA 2;
- Oracle's Elixir (https://oracleselixir.com) for League of Legends;
- Winston's Lab (www.winstonslab.com), OverBuff (www.overbuff.com), and Overwatch Tracker Network (https://overwatchtracker.com) for Overwatch;
- PUBG Stats (https://pubgstats.com) for PlayerUnknown's Battlegrounds;
- HSReplay.net (https://hsreplay.net) for Hearthstone;
- HOTS Logs (www.hotslogs.com) for Heroes of the Storm;

---

[8] See https://github.com/esports-bits/lol-data-solution
[9] See https://cod.tracker.gg/site-api
[10] See https://fortnitetracker.com/site-api
[11] See https://fortniteapi.com
[12] See https://overwatchy.com/docs
[13] See https://hearthstoneapi.com
[14] See https://hotsapi.net
[15] In addition, some third-party companies (e.g., ESPORT API, PandaScore) offer API-based data access as a paid service.

- Call of Duty Tracker Network (https://cod.tracker.gg); and
- Fortnite Tracker Network (https://fortnitetracker.com) and FortBuff (www.fortbuff.com) for Fortnite.

Many of these sites also offer data exploration interfaces that allow users to select desired observations and attributes from drop-down menus. In addition, and importantly for the purpose of this article, the preset and dynamic views available from these sites can usually be downloaded as CSV files. Indeed, the authors of this article recommend this approach to e-sports data collection to instructors and students who want to conduct analyses right away before learning how to write API-based queries or how to parse replay files.

Replay File Repositories

For all major e-sports games, each player's moves, actions, and in-game messages are recorded in a game log. Each of these events is digitally encoded in a way that allows users to (re)watch the match in its entirety through a compatible viewing program. For example, League of Legends "replay files" can be viewed through the ROFL Viewer.[16] After retrieving the replay files for the matches they have played in – files that are stored locally – users can fast-forward, "rewind," and pause the match footage as desired. Indeed, they can even choose different perspectives from which to view the footage.

The digital encoding of match events also allows analysts to parse the replay file (which is the totality of all encoded events) in order to extract numerous attributes (Eggert et al., 2015). Unfortunately, though, very few game publishers provide access to a central repository of replay files. In response to the demand by analysts for such files, though – demand that is driven by their wealth of available attributes – a small number of third-party sites have begun to crowdsource replay files and offer them publicly (see, e.g., HOTS Logs[17] and Fortnite Tracker[18]). OpenDota (www.opendota.com) is an exception in that it retrieves replay files directly from the game publisher (Valve) in order to provide a directory of pre-parsed DOTA 2 replay files.

As noted, though, unparsed replay files must be parsed before they can be analyzed. Unfortunately, parsing knowledge is gained along a fairly steep learning curve. For those who still want to parse replay files themselves, parsing libraries have been created as an aid for some e-sports games. For example, DOTA 2 parsing libraries include Clarity (Java), Alice (C++), Yasha (Go), and Smoke (Python).

CONTROLLING FOR THREE VARIABLES DURING DATA COLLECTION

When an e-sports analyst collects data to be analyzed, she must take into account certain variables. If she does not do so, then she may include observations that should have been excluded, which in turn may jeopardize the validity of results. Three such variables are discussed here in turn – namely, the game patch, skill level, and game mode – together with the reasons why analysts control for them.

[16] See https://sourceforge.net/projects/rofl-viewer
[17] See www.hotslogs.com/Replays/ReplaySearch for Heroes of the Storm
[18] See https://replays.fortnitetracker.com for Fortnite

Game Patch

In short, a "patch" is a major or minor update to an e-sports game. A patch may serve to fix known bugs and exploits, but its importance to performance analytics stems from another key purpose, namely, to "rebalance" the strengths and weaknesses of existing characters and items and/or to introduce new characters and items. These patches are seen as necessary because (1) one or more characters may have become (since the last patch) too strong or weak relative to other characters and/or (2) the game publisher has determined, usually based in part on feedback from players and fans, that a particular game strategy has become too effective (and thus used by nearly every team) or that existing gameplay has become too predictable and could thus benefit from new conditions and strategies. Indeed, the e-sports community uses the term "meta" to describe the set of characters, or the composition of characters and/or roles on a team, that is widely seen as consistently successful in professional play at a point in time (Liang, 2017). Patches therefore alter the "meta game" (Hodge et al., 2017), meaning that coaches and players must reconsider and possibly revise their strategies every time a major patch is introduced.

A patch also may introduce changes to (1) the capabilities and limitations of items purchased during a match, (2) the affordances and constraints of game maps, and/or (3) the mechanics of gameplay. An e-sports competitor may have to modify her gameplay and strategies accordingly each time a new patch is introduced.

As an example, the "meta" for the professional Overwatch League (OWL), at the time of this writing, is based on a team composition of three "tank" roles and three "support" (or healer) roles. No "damage" role is included in this composition. Moreover, many if not most of the teams employing this composition use the same six heroes: D.Va (tank); Reinhardt (tank); Zarya (tank); Brigitte (support); Zenyatta (support); and Lúcio (support). It is regarded as a safe and defensive "meta," but many fans find it predictable and somewhat dull to watch. As a result, some fans, and even some professional players, are reportedly hoping that Blizzard releases a patch that will make this "meta" less viable.

Patches may be frequent, depending on the publisher. From 2 January through 21 May 2019, for example, Valve introduced 24 minor updates, three minor patches (i.e., versions 7.21b, 7.21c, and 7.21d) and one major patch (i.e., version 7.21).[19] Major patches can have a significant impact on the "meta." As observed in one study, six major, consecutive DOTA 2 patches "changed 84 heroes (out of a pool of 111 heroes) and 39 items" (Summerville et al., 2016, p. 104).

So why must e-sports analysts control for "patch" as a variable? In short, if there are too many changes to characters, items, and gameplay from one point in time to another, then one can conclude that the unit of analysis (e.g., team, player, character) experienced very different gameplay conditions at the start of that time period versus at the end of it. As a result, findings that generalize across observations may be ambiguous or inconclusive when in fact they might have been conclusive if attributes had been limited to a single major patch, single season, or single "split" of a season.

Of course, this reasoning raises the question of how much change (to characters, items, maps, etc.) is too much change. While there is no consensus on an absolute threshold, Summerville et al. (2016) proposed that limiting observations to a single major patch is acceptable. For their study, Summerville et al. (2016) collected data "from all 1,518 professional

[19] See https://dota2.gamepedia.com/Patches

DOTA 2 matches played between 5 October and 16 December 2015," a period of time spanning "almost the entirety of [major] patch 6.85."

Skill Level

For every major e-sports game, the gap in skill level between novices and average players is very large, as is the gap between average players and top professionals. Indeed, one study found that there are even important gameplay differences between the highest-skilled non-professional DOTA 2 players and professional DOTA 2 players (Hodge et al., 2017).[20]

With regard to DOTA 2, for example, every player from casual novice to elite professional has a Matchmaking Rating (MMR) which places them within a certain rank. At the lowest level is the rank of Herald (0 to 800 MMR), followed by Guardian (800 to 1600 MMR), Crusader (1600 to 2400 MMR), Archon (2400 to 3300 MMR), Legend (3300 to 4100 MMR), Ancient (4100 to 5000 MMR), Divine (5000 to 5800 MMR), and Immortal (>5800 MMR).[21] (The numeric thresholds are approximate.) Each of these ranks comprises seven sub-ranks. As noted by Brown (2018), "the main purpose of MMR is to find and pair equal opponents and teammates for fair play." Similar ranking systems are employed in every major e-sports game.

E-sports analysts should control for the skill level of players and teams when conducting performance-based analyses (Yang et al., 2016). As both Drachen et al. (2014) and Hodge et al. (2017) have suggested, players at different skill levels employ contrasting strategies, respond differently to emergent conditions, and are dissimilar in their ability to execute, particularly when a tactic must be executed within seconds. For these reasons, the observations in a dataset should be limited according to rank or skill level (e.g., the Archon rank or MMR range in DOTA 2) or to a certain professional league or circuit.

Game Mode

While every major e-sports game offers game modes for unranked competitions and for practice against auto-generated opponents ("AI"), e-sports performance analyses are concerned with ranked competitions, or "ranked matchmaking" as it is usually called. For DOTA 2, ranked matchmaking is unlocked for any player at level 20; for League of Legends, it is unlocked at level 30.

For both DOTA 2 and League of Legends, ranked matchmaking comprises multiple game modes, with each mode centered on a different process through which teams ban and pick characters. For DOTA 2, for example, ranked matchmaking comprises three modes: Ranked All Pick; Captain's Mode; and Random Draft. Captain's Mode is the standard for most competitions, and is described briefly in the section (above) on intelligence needs. Controlling for game mode is important because teams and players that compete through different modes may employ different strategies and/or compete at different skill levels.

While controlling for these three variables promotes the validity of results, doing so may also create a challenge. Specifically, the number of pertinent observations that one is able to collect – of teams, players, characters, or items – may be limited. For example, there are only so

---

[20] Hodge et al. (2017) found that high-skill public DOTA 2 match data can be used with "slightly reduced accuracy" to predict winners in professional DOTA 2 matches.

[21] See https://dota2.gamepedia.com/Matchmaking/Seasonal_Rankings. In April 2019, a DOTA 2 player named coL.limmp had a solo MMR of 9402.

many teams that compete in a given league or circuit, and only so many characters or items associated with a given game. And when one controls for these three variables – patch, skill level, and game mode – it will likely have the effect of further reducing the number of available observations.

At the same time, most analytic methods and algorithms require a sufficient number of observations for valid results. Satisfying this condition is made more difficult (1) as the number of attributes increases (because observations must then increase at a greater rate) and/or (2) when supervised learning methods (e.g., decision trees, multiple linear regression) are used, given their need for training, test, and (possibly) validation datasets (Hodge et al., 2017). Of course, this challenge is certainly not unique to e-sports analytics, and various techniques and methods are available to address it (e.g., bootstrap sampling, *k*-fold cross-validation). A discussion of these techniques and methods lies outside the scope of this article, but can be found in most analytics and machine learning textbooks.

DATA ANALYTICS

In order to address intelligence needs around e-sports performance, the e-sports analyst must analyze the data she and others have collected. But what does e-sports analytics entail, exactly? What kind of analytic work is performed by an e-sports analyst? To answer this question, and in turn guide the planning of student research project and the development of lesson plans, two of this article's authors identified and analyzed 22 postings for e-sports analyst jobs.[22] The analysis of these job postings revealed that e-sports analytics mostly entails two broad areas of analytics work: (1) attribute creation for visualizations and dashboards; and (2) machine learning modeling. Each of these areas of work is discussed here in turn, followed by a brief discussion of the choice of programming languages and software applications for these two areas.

Attribute Creation for Visualizations and Dashboards

While analyzing the 22 postings for e-sports analytics jobs, it became clear that e-sports analytics work (like analytics work in most domains) partly entails the development of so-called "dashboards" and the visualizations that support them. For example, job postings from Riot Games called for applicants who can "develop new visualization dashboards for key metrics," "set up Tableau dashboards," and "share results in clear, illustrative dashboards." A job posting from Blizzard referred to the need to "tell stories with data and visualizations."

---

[22] In order to identify postings for e-sports analytics jobs, two of this article's authors examined the employment web pages for five major e-sports game publishers: Activision/Blizzard; Epic Games; PUBG Corporation; Riot Games; and Valve Software. Based on this examination, 19 pertinent jobs were identified in January 2019, including 11 from Riot Games, six from Activision/Blizzard, one from Epic Games, and one from Valve Software. The authors also searched for pertinent job postings on the web sites of major e-sports teams and major providers of e-sports research services. Based on these search efforts, pertinent job postings by one major e-sports team (100 Thieves) and two major e-sports research firms (MOBAlytics.gg and PandaScore) were identified.

The two authors then engaged in an inductive, qualitative analysis of the 22 job postings (Eisenhardt et al., 2016). Specifically, every skill, qualification, or responsibility mentioned in each job posting was labeled (coded) by each author. Labeling differences were discussed until consensus was reached. After a third iteration of labeling, the authors opted for a classification scheme with two top-level categories (i.e., attribute creation for visualizations and dashboards, machine learning modeling), and corresponding child categories.

E-sports analytics work thus involves the provision and visual presentation of salient descriptive statistics (attributes) to decision makers, communicated in a timely and effective way (Schubert et al., 2016). Given this dependence on attributes, it may be useful to identify some examples of basic and advanced attributes associated with e-sports performance. To this end, some of the more common team-based attributes associated with DOTA 2 are considered here. Drawing from team-based views from three third-party providers of DOTA 2 statistics,[23] the following common count-based attributes were identified, with each count corresponding to a certain match: kills; assists; deaths; gold gained, experience (XP) gained; and last hits (i.e., killing blows of an opponent creep or neutral creep).

Each of these counts (or totals) for a certain match also can be expressed as a rate and/or as a differential. For example, kills can be expressed as kills per minute (rate) or as "kill differential" (i.e., the number of opponents' kills subtracted from the number of kills by the focal team).[24] Expressing counts as rates or differentials (or averages, percentages, etc.) is an important analytic task because these computed attributes are usually more effective than counts in terms of correlating strongly with the target (response) attribute in predictive models (Kelleher and Tierney, 2018; Domingos, 2012). Just as importantly, they more accurately characterize an instance of the unit of analysis (e.g., a team and its effectiveness at killing) vis-a-vis other instances, and thus are more useful in visualizations and dashboards.

Moreover, if and when an analyst has a need to analyze teams over the course of many matches (e.g., a season-based analysis), she can calculate a per-match average for each of these measures (e.g., average kills per minute, average kill differential, etc.) Other multi-match, team-based attributes include win percentage (often referred to as "win rate") and Elo rating, which aims at summarizing a team's overall prowess by weighting more heavily (1) victories against teams with better win-loss records and (2) losses against teams with worse win-loss records.

Of course, many attributes are much more sophisticated than the examples provided above. Indeed, even count-based attributes can entail a good deal of time and effort. For example, Yang et al. (2014) wrote queries against DOTA 2 replay files in order to create several new count-based attributes, including one for the number of times a player casts a "detrimental" spell during a match. With regard to more sophisticated computed attributes, Van Allen (2018) has described how Doug Watson, who currently serves as Head of E-Sports Insights for Riot Games, created a "jungle proximity" attribute which helps predict team success in League of Legends:

> "The 'jungle' position splits [its] time between killing ambient wildlife in the 'jungle' for gold and experience, and trying to kill or hinder enemies on the main 'lanes'. In practice, the 'jungler' is a roaming source of pressure that often dictates their team's focus and pace... Watson's team analyzes the proximity of the jungler to each lane throughout the course of the game... [H]is team tracks each champion's X and Y coordinates on the map each second. Using their exact positions over a stretch of time, they can determine their proximity to each other and begin to analyze that connection."

As this anecdote suggests, the creation of advanced attributes depends heavily on the analyst possessing e-sports domain knowledge. In this regard, e-sports analytics is no different than analytics in most other domains; as Mitchell (1999, p. 33) observed, the analytic process

[23] The three third-party sites for DOTA 2 statistics that were reviewed for this purpose include OpenDota (www.opendota.com), DatDota (www.datdota.com), and DotaBuff (www.dotabuff).

[24] One of the most common attributes in MOBA-based games and first-person shooter (FPS) games is KDA, which is calculated as (kills+assists)/deaths.

requires "the use of human expert knowledge to formulate input variables." Along these lines, the authors' analysis of pertinent job postings found that three employers explicitly communicated that applicants should possess e-sports domain knowledge:

- "You need to game seriously and understand the vernacular" (MOBAlytics.gg)
- "You're a core gamer who takes play seriously" (Riot Games)
- "You use your gaming experience to interpret the data you're seeing (Blizzard)

It is also important to add that it is not enough, when creating attributes, to conclude that the unit of analysis is simply the team, player, item, character, or role. One must also know the associated unit of performance. Here, "unit of performance" refers to the match, round, encounter (Schubert et al., 2016), season, tournament, or other performance occurrence. For example, one common unit of analysis is team-in-match, meaning that every attribute is associated with a certain team's performance within a certain match. Another example is the player-in-season, meaning that every attribute is associated with a certain player's performance over the course of a certain season.

Machine Learning Modeling

The authors' analysis of pertinent job postings also revealed that contemporary e-sports analytics work entails the building of machine learning models. Indeed, every analyzed job posting called for applicants who can develop such models, including models that:

- Evaluate opponents' strengths and weaknesses (PandaScore, MOBAlytics.gg)
- Predict how changes to League of Legends champions (via new patches) will affect their win rates (100 Thieves)
- Recommend the items that a team should purchase during a match (100 Thieves)
- Minimize match queue times (Riot Games)
- Predict customer behavior (Valve Software)
- Determine customer lifetime value (Activision, Riot Games)

Machine learning is mostly synonymous with data mining and predictive analytics (Mitchell, 1999), and involves the use of a learning algorithm (e.g., linear regression, decision tree, *k*-nearest neighbor) to produce a mathematical formula (or model) which predicts output values based on a set of input values (Domingos, 2012).[25] (Unsupervised learning algorithms, such as *k*-means clustering, discover latent patterns among input attributes.) The model is "trained" using an existing dataset in which output values are known; its ability to then generalize beyond these training observations is assessed by seeing how well it predicts withheld values in a "test" dataset.

For example, consider an e-sports analyst who wants to maximize the rate at which she correctly predicts the win-loss percentage for any given team in a season of the North America League of Legends Championship Series (NA LCS). To do so, she creates a dataset in which team-in-season is the unit of analysis. Thus, each observation is specific to an NA LCS team (e.g., Cloud9, Echo Fox) and a certain season. Each attribute, then, represents a distinct characteristic of the team's performance throughout that season, expressed as a count (e.g., total

[25] More specifically, a machine learning algorithm optimizes similarities in a feature space (e.g., *k*-nearest neighbor), information gain (e.g., decision trees), event probability (e.g., naive Bayes classifier), or error reduction (e.g., multiple linear regression) (Kelleher et al., 2015).

number of last hits), a differential (e.g., total number of opponents' deaths subtracted from total number of focal team's deaths), a per-game average (e.g., net worth per game, XP/minute per game), a percentage (e.g., percentage of matches with first kill), or some other calculation (e.g., KDA per game). The dataset also includes a target (response) attribute – in this case win-loss percentage – that the analyst hopes to predict as accurately as possible. To accomplish this, she draws from her domain knowledge, and possibly from knowledge possessed by other domain experts, to select certain attributes to include in the model (Domingos, 2012). In the dataset that trains the model, the values for win-loss percentage are included; in the dataset that tests the model, they are withheld. After the model has been applied to the test data, the analyst is able to gauge the model's success (e.g., percentage of correct answers).[26] The analyst then deploys the model with the set of attributes that best predicts the target attribute.

The authors' analysis of existing research found that many machine learning methods and algorithms have been used to investigate e-sports performance. Examples include:

- Association rule learning (Looi et al., forthcoming);
- Cluster analysis (Looi et al., forthcoming);
- Decision trees, including boosted decision trees (Semenov et al., 2017; Yang et al., 2014) and random forests (Eggert et al., 2015; Kinkade et al., 2015);
- Logistic regression (Looi et al., forthcoming; Hodge et al., 2017; Semenov et al., 2017; Yang et al., 2016; Eggert et al., 2015; Kinkade et al., 2015);
- Multiple linear regression (Andersson and Alin, 2018);
- Naive Bayes classifier (Semenov et al., 2017; Eggert et al., 2015) and Bayesian nets (Summerville et al., 2016);
- Neural networks (Batsford, 2014), including long short-term memory recurrent neural networks (Summerville et al., 2016); and
- Support vector machines (Eggert et al., 2015).

The Choice of Programming Languages and Software Applications for E-Sports Analytics Work

As one tech-writer observed, "there are a seemingly unending number of coding languages, frameworks, and tools" that can be used for basic and advanced analytics (Goodman, 2019). Given this abundance of languages and software applications, which ones should be used for e-sports analytics? The authors' analysis of 22 pertinent job postings may help in answering this question. With regard to programming languages, SQL (15 job postings), Python (10), and R (5) were mentioned most frequently. Java, C++, PHP, and various code libraries for machine learning (e.g., TensorFlow, Keras, PyTorch) were mentioned in multiple postings.

These findings are mostly reflected in a recent survey of respondents who identify as data scientists (Suda, 2018). The results of this survey indicate that data scientists use SQL (64% of respondents), Python (63%), and R (54%) the most. In addition, more than 50 percent of respondents use a machine learning library, with Scikit-learn (37%) being the most popular.

With regard to software applications, the authors' analysis of job postings revealed seven broad types of applications used by e-sports employers:

- Spreadsheets, including MS Excel (mentioned in 8 job postings) and Looker (1);

[26] Our analysis of pertinent job postings found that multiple employers called for applicants who are able to interpret model diagnostics, evaluate model outputs, and revise the model accordingly.

- Visualization, including Tableau (9) and Kibana (1);
- Statistical, including Matlab (2), SPSS (2), SAS (1), and Systat (1);
- Databases, including MySQL (2), Amazon Redshift (1), and Apache Cassandra (1);
- Platform middleware, including Google Cloud (1), Apache Hadoop, (1), Apache Hive (1), and Apache Spark (1);
- Search engines, including ElasticSearch (1);
- Presentation, including MS Powerpoint (4); and
- Project management, including Trello (1) and Basecamp (1).

The results of Suda's (2018) survey of data scientists reflected these findings as well, with respondents using MS Excel (67%), Tableau (32%), MySQL (37%), and Apache Hadoop (18%) frequently relative to most other applications in their category.

Of course, a single e-sports course, module, or research project would likely only be able to focus on a very small number of programming languages and software applications. With this in mind, Goodman (2019) argues that analysts "don't have to know every language and tool, but should go deep on the basic tools [they] use daily." Goodman echoes this article's findings by recommending SQL, R (and its ggplot package), Python, and three Python libraries (i.e., Pandas, NumPy, and SciPy) as the must-know basic tools.

CONCLUSION

As the global e-sports industry continues to expand at a rapid pace, the e-sports labor market expands commensurately. A growing portion of this e-sports labor market comprises e-sports analytics jobs which help address intelligence needs around e-sports performance. This article described three such needs in detail: team performance and strategy; player strategy; and character and item strategy. Next, this article provided practical and detailed information about two key concerns related to e-sports data collection. With regard to the first concern, three common types of data sources for e-sports games were detailed: application programming interfaces (APIs); third-party APIs and statistics providers; and replay file repositories. With regard to the second concern, the authors described three variables that an analyst should control for during data collection: the game patch; skill level; and game mode.

In the article's fourth and final section, the authors drew from an analysis of recent postings for e-sports analytics jobs to identify and examine two broad areas of e-sports analytics work: attribute creation for visualizations and dashboards; and machine learning modeling. The article concluded with a brief discussion of popular programming languages and software applications for e-sports analytics. Through these four sections, the authors' intent was to provide instructors and students with an overview of the e-sports industry, a justification for resources earmarked for instruction around e-sports performance analytics, and, most importantly, detailed discussions of certain key areas that constitute e-sports performance analytics.

REFERENCES

Andersson, J. and Alin, O. (2018). Analyzing Factors Contributing to the Success of a Team in DOTA 2. Thesis for degree, KTH Royal Institute of Technology. Accessed in March 2019 from http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1216849&dswid=-5549.

Baccellieri, E. (2018). Twitch signs two-year deal with Overwatch League worth at least $90 million. *Compete*, 9 January 2018. Accessed in March 2019 from https://compete.kotaku.com/report-twitch-signs-two-year-deal-with-overwatch-leagu-1821932438.

Bamforth, E. (2018). The University of Akron says it has the most esports-dedicated space of any school worldwide. *Cleveland.com*, 16 August 2018. Accessed in January 2019 from https://www.cleveland.com/metro/index.ssf/2018/08/university_of_akron_to_dedicat.html.

Batsford, T. (2014). Calculating optimal jungling routes in DOTA 2 using neural networks and genetic algorithms. *Game Behaviour*, 1(1).

Bozorgzadeh, A. (2017). Machine learning will be a game-changer for eSports. *Venture Beat*, 11 March 2017. Accessed in January 2019 from https://venturebeat.com/2017/03/11/machine-learning-will-be-a-game-changer-for-esports/.

Brown, M. (2018). DOTA 2 ranks and ranking system. *DMarket*, 28 November 2018. Accessed in May 2019 from https://dmarket.com/blog/dota2-ranks.

Carp, S. (2018). Syracuse University joins forces with Twitch for esports course. *SportsPro*, 8 June 2018. Accessed in March 2019 from www.sportspromedia.com/news/twitch-esports-college-course-syracuse-university.

Davenport, T.H. (2014). What businesses can learn from sports analytics. *MIT Sloan Management Review*, 55(4), 10-13.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 61(11), 27-29.

Drachen, A., Yancey, M., Maguire, J., Chu, D., Wang, Y.I., Mahlman, T., Schubert, M., and Klabjan, D. (2014). Skill-based differences in spatio-temporal team behavior in Defence of the Ancient 2. *Proceedings of the 2014 IEEE Games Media Entertainment Conference*, 22-24 October 2014, Toronto, Canada.

Drachen, A. (2017). The case for data in eSports. Game Analytics Resources v. Anders Drachen, 9 August 2017. Accessed in January 2019 from https://andersdrachen.com/2017/08/09/the-case-for-data-in-esports/.

Eggert, C., Herrlich, M., Smeddinck, J., and Malaka, R. (2015). Classification of player roles in the team-based multi-player game DOTA 2. *Proceedings of the 14th International Conference on Entertainment Computing* (ICEC), September 2015, Trondheim, Norway.

Eisenhardt, K.M., Graebner, M.E., and Sonenshein, S. (2016). Grand challenges and inductive methods: rigor without rigor mortis. *Academy of Management Journal*, 59(4), 1113-1123.

Gao, L., Judd, J., Wong, D., and Lowder, J. (2013). Classifying DOTA 2 hero characters based on play style and performance. Course report, University of Utah, 2013.

Goldman Sachs Investment Research. (2019). eSports joins the big leagues. *Goldman Sachs Insights*, January 2019. Accessed in February 2019 from https://www.goldmansachs.com/insights/infographics/e-sports.

Goodman, J. (2019). 12 things I wish I'd known before starting as a data scientist. *Medium*, 20 February 2019. Accessed in March 2019 from https://medium.com/deliberate-data-science/12-things-i-wish-id-known-before-starting-as-a-data-scientist-45989be6300e.

Hodge, V., Devlin, S., Sephton, N., Block, F., Drachen, A., and Cowling, P. (2017). Win prediction in eSports: mixed-rank match prediction in multi-player online battle arena games. Accessed in February 2019 from https://arxiv.org/abs/1711.06498.

Iqbal, M. (2019). Twitch revenues and usage statistics. *BusinessOfApps*, 27 February 2019. Accessed March 2019 from www.businessofapps.com/data/twitch-statistics.

Kelleher, J.D. and Tierney, B. (2018). *Data Science*. Cambridge, MA: MIT Press.

Kelleher, J.D., MacNamee, B., and D'Arcy, A. (2015). *Fundamentals of Machine Learning for Predictive Data Analytics*. Cambridge, MA: MIT Press.

Kinkade, N., Jolla, L., and Lim, K. (2015). DOTA 2 win prediction. Course report, University of California-San Diego, 2015.

Lam, S. (2017). eSports as a goldmine for data analytics. *ComputerWorldHK*, 17 October 2017. Accessed in February 2019 from https://www.cw.com.hk/data-management/esports-as-a-goldmine-for-data-analytics.

Liang, J. (2017). 'What's in a meta?' Analyzing champion viability in Esports. *Proceedings of the 2017 MIT Sloan Analytics Conference*, 3-4 March 2017, Boston, MA.

Looi, W., Dhaliwal, M., Alhajj, R., and Rokne, J. (Forthcoming.) Recommender systems for items in DOTA 2. *IEEE Transactions on Games*. DOI: 10.1109/TG.2018.2844121.

Mitrevski, L. (2018). The University of Akron debuts 5,000-square-foot esports facility. *ESI: The Esports Insider*, 5 October 2018. Accessed in January 2019 from https://esportsinsider.com/2018/10/university-of-akron-esports-facility/.

Mitchell, T.M. (1999). Machine learning and data mining. *Communications of the ACM*, 42(11), 31-36.

Newzoo. (2018). *2018 Global eSports Market Report*. Amsterdam, The Netherlands: Newzoo.

Niederhoffer, N. (2018). As Overwatch League team hires Winston Lab founder, eSports analytics get a huge upvote. *Washington Post*, 13 November 2018. Accessed in January 2019 from https://www.washingtonpost.com/sports/2018/11/13/overwatch-league-team-hires-winstons-lab-founder-esports-analytics-get-huge-upvote/.

Ong, H.Y., Deolalikar, S., and Penge, M.V. (2015). Player behavior and optimal team composition in online multiplayer games. Accessed in May 2019 from https://arxiv.org/abs/1503.02230.

Pannekeet, J. (2019). Global esports economy will top $1 billion for the first time in 2019. *Newzoo Insights*, 12 February 2019. Accessed in February 2019 from https://newzoo.com/insights/articles/newzoo-global-esports-economy-will-top-1-billion-for-the-first-time-in-2019/.

Pettit, E. (2018). U. of Akron will phase out 80 degree programs and open new esports facilities. The Chronicle of Higher Education, 16 August 2018. Accessed in January 2019 from https://www.chronicle.com/article/U-of-Akron-Will-Phase-Out-80/244293.

Schubert, M., Drachen, A., and Mahlmann, T. (2016). Esports analytics through encounter detection. *Proceedings of the 2016 MIT Sloan Analytics Conference*, 11-12 March 2017, Boston, MA.

Semenov, A., Romov, P., Korolev, S., Yashkov, D., Neklyudov, K. (2017). Performance of machine learning algorithms in predicting game outcome from drafts in DOTA 2. *Proceedings of the International Conference on Analysis of Images, Social Networks, and Texts* (pp. 26-37), April 2016, Yekaterinburg, Russia.

Suda, B. (2018). 2017 *Data Science Salary Survey*. Sebastopol, CA: O'Reilly Media, Inc.

Summerville, A., Cook, M., Steenhuisen, B. (2016). Draft analysis of the ancients: predicting draft picks in DotA 2 using machine learning. *Proceedings of the 12th Artificial Intelligence and Interactive Digital Entertainment Workshop*, 8 October 2016, Burlingame, CA.

TwitchTracker. (2019). Twitch game statistics. TwitchTracker, 22 March 2019. Accessed in March 2019 from https://twitchtracker.com/statistics/games.

Van Allen, E. (2018). Big data is the future of eSports. *Kotaku.com*, 29 June 2018. Accessed in January 2019 from https://compete.kotaku.com/big-data-is-the-future-of-esports-1827206481.

Yang, P., Harrison, B., and Roberts, D.L. (2014). Identifying patterns in combat that are predictive of success in MOBA games. Proceedings of Foundations of Digital Games 2014, 3-7 April 2014, Ft. Lauderdale FL.

Yang, Y., Qin, T., and Lei, Y. (2016). Real-time eSports match result prediction. Accessed in April 2019 from https://arxiv.org/abs/1701.03162.

APPENDIX

Figure 1: An Excerpt of JSON-Based Results from an API Query for Attributes Associated with Heroes in DOTA 2.


```
[{"id":1,"name":"npc_dota_hero_antimage","localized_name":"Anti-Mage","primary_attr":"agi","attack_type":"Melee","roles":["Carry","Escape","Nuker"],"legs":2},
{"id":2,"name":"npc_dota_hero_axe","localized_name":"Axe","primary_attr":"str","attack_type":"Melee","roles":["Initiator","Durable","Disabler","Jungler"],"legs":2},
{"id":3,"name":"npc_dota_hero_bane","localized_name":"Bane","primary_attr":"int","attack_type":"Ranged","roles":["Support","Disabler","Nuker","Durable"],"legs":4},
{"id":4,"name":"npc_dota_hero_bloodseeker","localized_name":"Bloodseeker","primary_attr":"agi","attack_type":"Melee","roles":["Carry","Disabler","Jungler","Nuker","Initiator"],"legs":2},
{"id":5,"name":"npc_dota_hero_crystal_maiden","localized_name":"Crystal Maiden","primary_attr":"int","attack_type":"Ranged","roles":["Support","Disabler","Nuker","Jungler"],"legs":2},
{"id":6,"name":"npc_dota_hero_drow_ranger","localized_name":"Drow Ranger","primary_attr":"agi","attack_type":"Ranged","roles":["Carry","Disabler","Pusher"],"legs":2},
{"id":7,"name":"npc_dota_hero_earthshaker","localized_name":"Earthshaker","primary_attr":"str","attack_type":"Melee","roles":["Support","Initiator","Disabler","Nuker"],"legs":2},
{"id":8,"name":"npc_dota_hero_juggernaut","localized_name":"Juggernaut","primary_attr":"agi","attack_type":"Melee","roles":["Carry","Pusher","Escape"],"legs":2},
{"id":9,"name":"npc_dota_hero_mirana","localized_name":"Mirana","primary_attr":"agi","attack_type":"Ranged","roles":["Carry","Support","Escape","Nuker","Disabler"],"legs":2},
{"id":10,"name":"npc_dota_hero_morphling","localized_name":"Morphling","primary_attr":"agi","attack_type":"Ranged","roles":["Carry","Escape","Durable","Nuker","Disabler"],"legs":0},
{"id":11,"name":"npc_dota_hero_nevermore","localized_name":"Shadow Fiend","primary_attr":"agi","attack_type":"Ranged","roles":["Carry","Nuker"],"legs":0},
{"id":12,"name":"npc_dota_hero_phantom_lancer","localized_name":"Phantom Lancer","primary_attr":"agi","attack_type":"Melee","roles":["Carry","Escape","Pusher","Nuker"],"legs":2},
{"id":13,"name":"npc_dota_hero_puck","localized_name":"Puck","primary_attr":"int","attack_type":"Ranged","roles":["Initiator","Disabler","Escape","Nuker"],"legs":2},
{"id":14,"name":"npc_dota_hero_pudge","localized_name":"Pudge","primary_attr":"str","attack_type":"Melee","roles":["Disabler","Initiator","Durable","Nuker"],"legs":2},
{"id":15,"name":"npc_dota_hero_razor","localized_name":"Razor","primary_attr":"agi","attack_type":"Ranged","roles":["Carry","Durable","Nuker","Pusher"],"legs":0},
{"id":16,"name":"npc_dota_hero_sand_king","localized_name":"Sand King","primary_attr":"str","attack_type":"Melee","roles":["Initiator","Disabler","Support","Nuker","Escape","Jungler"],"legs":6},
{"id":17,"name":"npc_dota_hero_storm_spirit","localized_name":"Storm Spirit","primary_attr":"int","attack_type":"Ranged","roles":["Carry","Escape","Nuker","Initiator","Disabler"],"legs":2},
{"id":18,"name":"npc_dota_hero_sven","localized_name":"Sven","primary_attr":"str","attack_type":"Melee","roles":["Carry","Disabler","Initiator","Durable","Nuker"],"legs":2},
{"id":19,"name":"npc_dota_hero_tiny","localized_name":"Tiny","primary_attr":"str","attack_type":"Melee","roles":["Carry","Nuker","Pusher","Initiator","Durable","Disabler"],"legs":2},
{"id":20,"name":"npc_dota_hero_vengefulspirit","localized_name":"Vengeful Spirit","primary_attr":"agi","attack_type":"Ranged","roles":["Support","Initiator","Disabler","Nuker","Escape"],"legs":2},
```


Figure 2: An Excerpt of a View Containing Statistics for Teams that Competed in the 2019 Regular Season of the North American League of Legends Championship Series (NA LCS). Source: Oracle's Elixir (https://oracleselixir.com), 28 May 2019.

TEAM	GP	W	L	AGT	K	D	K:D	CKPM	GPR	GSPD	EGR	MLR	GD@15	FB%	FT%	F3T%
100 Thieves	18	4	14	35.9	137	200	0.69	0.52	-0.49	-4.4%	42.8	-20.6	-558	33%	44%	33%
Cloud9	18	14	4	33.8	253	165	1.53	0.69	0.66	6.8%	57.3	20.5	674	61%	67%	72%
Clutch Gaming	18	5	13	36.2	182	232	0.78	0.64	-0.44	-0.9%	44.6	-16.8	-415	67%	39%	50%
Counter Logic Gaming	18	7	11	36.3	135	183	0.74	0.49	-0.78	-5.5%	39.6	-0.7	-882	56%	39%	39%
Echo Fox	18	8	10	35.1	184	194	0.95	0.60	-0.20	-1.1%	51.4	-6.9	21	28%	56%	33%
FlyQuest	19	10	9	36.0	202	192	1.05	0.58	0.38	1.9%	52.7	0.0	340	58%	42%	42%
Golden Guardians	19	9	10	35.7	195	219	0.89	0.61	0.18	-2.0%	55.3	-7.9	291	53%	53%	47%
OpTic Gaming	18	7	11	35.6	162	218	0.74	0.59	-0.26	-3.2%	47.0	-8.1	-331	61%	39%	56%
Team Liquid	18	14	4	35.5	235	131	1.79	0.57	0.72	4.8%	56.3	21.5	271	33%	61%	72%
Team SoloMid	18	13	5	38.2	214	167	1.28	0.55	0.24	3.6%	52.8	19.5	554	50%	61%	56%