Proceedings, The Thirteenth AAAI Conference on
Artificial Intelligence and Interactive Digital Entertainment (AIIDE-17)

Predicting Victory in a Hybrid Online Competitive Game: The Case of *Destiny*

Yaser Norouzzadeh Ravari,[1] Pieter Spronck,[1] Rafet Sifa,[2] Anders Drachen[3]
[1] Tilburg Center for Cognition and Communication, Tilburg University, Tilburg, The Netherlands
[2] Fraunhofer IAIS, Germany
[3] DC Labs, University of York, United Kingdom
Y.NorouzzadehRavari@uvt.nl, p.spronck@uvt.nl, Rafet.Sifa@iais.fraunhofer.de, anders.drachen@york.ac.uk

Abstract

Competitive multi-player game play is a common feature in major commercial titles, and has formed the foundation for esports. In this paper, the question whether it is possible to predict match outcomes in First Person Shooter-type multi-player competitive games with mixed genres is addressed. The case employed is *Destiny*, which forms a hybrid title combining Massively Multi-player Online Role-Playing game features and First-Person Shooter games. *Destiny* provides the opportunity to investigate prediction of the match outcome, as well as the influence of performance metrics on the match results in a hybrid multi-player major commercial title. Two groups of models are presented for predicting match results: One group predicts match results for each individual game mode and the other group predicts match results in general, without considering specific game modes. Models achieve a performance between 63% and 99% in terms of average precision, with a higher performance recorded for the models trained on specific multi-player game modes, of which *Destiny* has several. We also analyzed performance metrics and their influence for each model. The results show that many key shooter performance metrics such as Kill/Death ratio are relevant across game modes, but also that some performance metrics are mainly important for specific competitive game modes. The results indicate that reliable match prediction is possible in FPS-type esports games.

Introduction

This paper deals with match result prediction in the game *Destiny*. Victory prediction concerns analyses across electronic sports (esports), notably Multiplayer Online Battle Arena (MOBA) games like DotA2, League of Legends and StarCraft (Schubert, Drachen, and Mahlmann 2016). Esports is growing by the number of players, viewers, and funds. Esports revenues increased 41% in 2017 and Newzoo predicted it will reach 1.5 billion dollars in 2020. In 2017, the esports audiences will be around 384 million.

The goal of victory prediction is to predict which player or which team of players will win a match. The definition of victory depends on the game dynamics. Some matches are win-loss, while in other matches, the rank of players determines victory. Given the variety of game mechanics, there is a broad design space available for victory conditions in esports games and beyond. Victory prediction can be studied in human vs. Artificial Intelligence (AI), human vs. human, or AI vs. AI situations. The victory prediction can be done based on pre-match, within-match, or post-match features. In this work, we study victory prediction on human vs. human player matches using post-match features, for the purpose of predicting the outcome of a match.

Victory prediction is of direct interest to esports industry, audiences, and researchers. On the research side, esports provide a complex testing ground for machine learning algorithms, thanks to the availability of voluminous, varied and volatile datasets (Schubert, Drachen, and Mahlmann 2016; Yang, Harrison, and Roberts 2014). For example, *Destiny*'s back end servers contain more than 1,400 features per player character across millions of players. In *Destiny*, players can participate in Player versus Environment (PvE) and Player versus Player (PvP) activities. *Destiny* includes various game modes with different objectives within both types of activities. Here the focus is on the PvP game modes. Victory prediction in *Destiny* forms a unique case within the larger domain of esports analytics, as the game combines Massively Multiplayer Online Role-Playing Game (MMORPG), such as class selection, leveling and experience, and First Person Shooter (FPS) elements, such as a heavy emphasis on weapons and tactics. It also provides multiple varying competitive game modes to investigate at the same time. Furthermore, despite being a hybrid online title, the performance metrics important in *Destiny* are similar to other competitive FPS titles such as the major esports title *CounterStrike*, and to a degree also to MMORPGs such as *World of Warcraft*, making *Destiny* a broadly applicable case.

Prediction of match results has been studied in previous research for matches between AI players and for matches between humans and AI players (Bakkes, Spronck, and van den Herik 2007; Hsu, Hung, and Tsay 2013; Erickson and Buro 2014). Predicting the result of PvP matches is rare, though in early work Ravari, Bakkes, and Spronck (Ravari, Bakkes, and Spronck 2016) predicted match results in the Real-Time Strategy game *StarCraft*. The work presented here directly advances the state or the art by focusing on FPS situations, and also by operating across multiple variations of the competitive team-based form, which is typical in esports across FPS and Multi-player Online Battle Arena

(MOBA) titles (Schubert, Drachen, and Mahlmann 2016; Yang, Harrison, and Roberts 2014). The present work is the first study on match result prediction in *Destiny*. Our models focus on predicting match results across win-loss and ranking matches in 13 different PvP game modes. The findings presented distinguish metrics that are most important for determining match results in general, and some that are important for specific PvP modes in *Destiny*. This emphasizes the need for careful modeling across game modes in esports games in general.

In the following sections, we present related work, an overview of *Destiny* and of the dataset that we used, the features that we selected, the experimental setup, our results, and our conclusions.

Related Work

Match prediction in competitive games has been studied primarily from two perspectives: 1) AI-driven work in RTS games such as *StarCraft* for the purpose of developing AI players. For example, Cole, Louis, and Miles (2004) developed AI bots at the expert level for the FPS *CounterStrike*. 2) behaviorally driven work in esports for the purpose of providing knowledge to players and teams. For example, Schubert, Drachen, and Mahlmann (2016) developed encounter-based models for evaluating MOBA matches and predicting match results. Additionally, player behavior has been investigated from a broad array of perspectives across scientific disciplines. For example, Schatten, Tomičić, and Đurić (2015) proposed an agent-based model to study the organizational behavior of players. For reasons of space, the focus will here be on the references most directly relevant to the work presented here.

The prediction of match- and combat outcome or match results has been the focus of research across different genres of games, notably RTS games. For example Bakkes, Spronck, and van den Herik (2007) utilized match status in different phases to predict the match result in *SPRING*. Yang, Harrison, and Roberts (2014) investigated common patterns of winning teams in combat tactics. Erickson and Buro (2014) used players' features and battle information to predict match results in *StarCraft*. Ravari, Bakkes, and Spronck (2016) investigated winner prediction for all match types using within-match features of *StarCraft*, describing the most important skills involved in winning a match. In contrast to this research, the focus here is on post-match results prediction within and across multiple different game modes.

In esports, match prediction forms a key focus in the limited literature that is available, recently summarized by Schubert, Drachen, and Mahlmann (2016). While there has been limited to no work on match prediction in FPS games outside of the broader esports community, analytics for MOBAs has been the focus of more than a dozen publications. The consensus is that match prediction is possible but there is as yet no substantial body of publicly available work to compare performance results with (Schubert, Drachen, and Mahlmann 2016; Yang, Harrison, and Roberts 2014).

Destiny: gameplay

Destiny has a science fiction story that merges the characteristics of different game genres. It provides a wide range of PvE and PvP game modes. In *Destiny*, players can participate in missions, events and raids. They engage in combat and other activities to gain new abilities, more powerful guns, and to level up their character. Players can run, jump, crouch, shoot, and use melee weapons. *Destiny* includes three main character classes: Hunter, Titan, and Warlock. Each has different strengths and weaknesses, with access to various abilities. The player chooses one of these classes at the start of a match. Each class includes subclasses that determine the specific upgrades and improvements of the main class.

In *Destiny*, human Player versus (human) Player (PvP) content is accessed via the Crucible, which is a hub for PvP content in the game. In the Crucible, players can choose different modes of play, with varying rules and objectives. In most of the game modes levels and gear are disabled, meaning that bonuses conferred by these are equalized among the players. Weapon stats and abilities are generally enabled. Some of the game modes are only available during specific events.

Points are generally scored by killing (with bonuses for particular kinds of kills), and assisting and supporting team mates. In particular game modes points can also be scored for capturing or neutralizing zones, reviving team mates, and deploying or neutralizing probes. Points for particular actions may vary between game modes, and there is also variation in how many points go to the team, and how many points go to individual players. Most game modes represent win-loss matches, i.e., players win as a team. A few game modes are free-for-all matches, in which each player receives a rank at the end to compare his or her performance with the performances of the other players.

There are 13 PvP modes in *Destiny* at the time of writing: Skirmish, Control, Salvage, Clash, Trials of Osiris, Doubles, Iron Banner, Elminiation, Rift, Mayhem Clash, Zone Control, Rumble and Supremacy. For reasons of space not all of these are described here, but they include the following:

- Skirmish: Skirmish is a 3v3 PvP mode whereby the first team which earns 5,000 points wins the match. The objective is to keep the teammates alive and fight the enemy.
- Control: Control is a 6v6 PvP mode. In this mode, three flags are scattered around the map, and teams must capture flags and defend them.
- Salvage: Salvage is a 3v3 mode. The goal is to capture a target point and collect secrets. The team that did not capture the target point, must interrupt the first team. The team that collects more secrets in a limited time wins the match.
- Clash: Clash is a 6v6 PvP mode in which players team up in 2 teams. Teams fight to earn 10,000 points by getting kills and assists.

Dataset

The *Destiny* dataset includes players' end-match performances from September 2014 to January 2016. Each player can have up to three characters. In total, the dataset includes performances of about 15,000 characters and 9,000 players. The number of players in Titan, Hunter, and Warlock class types are 4,000, 5,000, and 5,000, respectively. The numbers of samples for Titan, Hunter, and Warlock players are 600,000, 800,000, and 700,000, respectively. Each sample shows a summary of the performance of a character of a player at the end of a match. This information includes the player Id, character Id, class type, date of activity, and more than 1,000 features that represent the player's performance metrics such as the number of kills, deaths, and assists.

In PvP game modes, two teams of players play against each other. The number of players in each team can vary. Players team up before the match. Unfortunately, in the dataset, team information and match Id are missing. This entails that we are unable to determine which players were in a match together.

In the dataset, the result of a match is denoted by a variable named 'standing.' Its value is an integer in the range 0 to 5. In win-loss matches, the match has a winning and a losing team; the standing value is 0 if the player was part of the winning team, and 1 if the player was part of the losing team. In free-for-all matches, the standing value is in the range 0 to 5, indicating the player's rank, with 0 going to the best player. In this dataset, we have 2 million samples for win-loss matches, and 145,000 samples for free-for-all matches.

Feature	mean	stdev
stat-agility	2.58	5.21
stat-armor	2.22	6.33
stat-discipline	90.96	149.51
stat-intellect	103.00	168.07
stat-light	76.93	249.50
stat-optics	21.87	64.28
stat-recovery	2.23	5.06
stat-strength	72.96	115.46
activity duration seconds	1504.63	601.82
activity length	668.21	638.78
assists	2.71	3.53
average score per kill	195.85	189.35
average score per life	257.73	186.96
class type	0.78	1.03
completed	0.27	0.91
completion reason	6.32	0.27
current progress	51189.37	316352.80
daily progress	1399.75	50.05
deaths	4.95	10.93
gender type	0.40	0.20
kills	6.63	10.76
kills deaths assists	1.29	1.32
kills deaths ratio	1.18	1.14
leave remaining seconds	25.79	0.39
level	5.03	37.20
minutes played this session	71.83	62.37
minutes played total	22459.69	28466.16
mode	4.34	13.13
next level at	4278.25	2726.71
percent to next level	21.54	8.60
player count	3.30	11.03
progress to next level	2075.96	807.86
weekly progress	5122.14	297.93

Table 1: Mean and standard deviation of the primary features in the PvP dataset from *Destiny* from (Ravari 2017).

Features

We used 34 features that were tracked in the game, and these represent typical FPS metrics as well as metrics that tries to capture the unique elements of *Destiny*, e.g., stats and assists. Table 1 shows the list of used features, and mean and standard deviation of them in our dataset. The explanation of some of the features is as follows:

- Stat-agility: affects movement speed and jumps height.
- Stat-armor: the higher Armor, less damage player will take.
- Stat-discipline: affects grenade cooldown time.
- Stat-intellect: influences the Super cooldown time.
- Stat-light: it is the second leveling method for players that reached maximum Level and increases output damage.
- Stat-optics: influences zoom capability of the weapon while aiming.
- Stat-recovery: shows how fast player's heath and shields regenerated after taking damage.
- Stat-strength: influences the cooldown of melee ability.
- Completion reason: multiple possibilities, a.o.: killing all of the opponents, earning specified points, reaching the match time limit, or achieving the objective.
- Current progress: earned points.
- Leave remaining seconds: remaining seconds of an activity, if a player leaves the activity before it ends.
- Next level at: required points to reach next level.
- Player count: the number of players in the match.

Experimental Setup

For our classification efforts, since win-loss matches have only two possible standing values, we formulated the prediction for win-loss matches as binary classification, while for free-for-all matches we used multiclass classification. The player's features are considered inputs, and the standing value the classifier's output. We use the one-vs-all strategy for multiclass classification, because this strategy is computationally efficient and interpretable. In this strategy, one classifier is fitting each class. Therefore, the number of classifiers is equal to the number of classes.

We employed two state-of-the-art classification methods: Gradient Boosting (GB) (Friedman 2002), and Random Forests (RF) (Breiman 2001). GB uses an ensemble of weak learners, such as regression trees, and optimizes a loss function to generalize them. GB is robust to outliers, can handle combined type features, does not need to normalize the in-

Model	Classification	AUC	avg precision
win-loss	GB-Binary	82%	81%
win-loss	RF-Binary	84%	84%
ranking	GB-Multiclass	88%	63%
ranking	RF-Multiclass	90%	68%
binary-ranking	GB-Binary	95%	94%
binary-ranking	RF-Binary	94%	94%

Table 2: Performance of combined models from (Ravari 2017).

puts, and can handle non-linear dependencies between the feature values and the outputs. RF also is an ensemble learning method, which uses decision trees for prediction. GB and RF have been used successfully for prediction tasks in video games (Sifa et al. 2015; Ravari, Bakkes, and Spronck 2016; Sifa et al. 2016; Drachen et al. 2016). For instance, Ravari, Bakkes, and Spronck (2016) successfully employed GB and RF to predict the match winner in different match types for *Starcraft* and Sifa et al. (2016) used RF to find how much spatio-temporal features affect the retention prediction.

Here two groups of models were developed: *combined models* and *individual models*. Combined models predict the match result by ignoring the game modes, while individual models take the game mode into account. We distinguish the following three types of combined models:

- Combined models:
  - Win-loss model: this model predicts the result of win-loss game modes (0 or 1).
  - Ranking model: this model predicts the rank of player for free-for-all game modes (range 0 to 5).
  - Binary-ranking model: this model is a binary version of the ranking model. In this model, we divided the ranks into two groups: the first group includes ranks 0, 1, and 2, while the second group includes ranks 3, 4, and 5.
- Individual models: 13 models predict the match results for each game modes (binary or multi-class, depending on the game mode).

To train the models, we divided data into randomized training (70%) and test (30%) sets, ensuring that a player who is in the training set is not in the test set. For each model, we trained the model on the training, and we evaluated the model on the test set.

Results

In this section, we present the performance of our models. We also discuss the top-5 features for each model. The performance of the models are represented by Area Under Curve (AUC) and average precision. These are two common metrics to show the performance of classifiers in machine learning (Fogarty, Baker, and Hudson 2005; Friedman, Hastie, and Tibshirani 2001).

AUC is the area under the Receiver Operating Characteristic (ROC) curve. ROC curve represents the true positive rate (recall) against the false positive rate (FPR) for different classification thresholds. Generally, AUC is in [0.5, 1]. AUC=1 means ideal performance, while AUC=0.5 represents the worst performance.

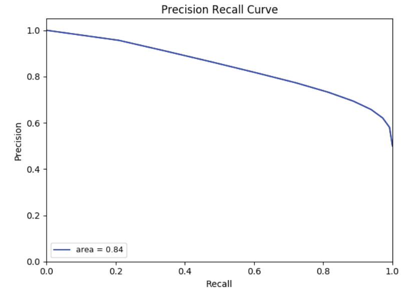

Figure 1: Precision-recall curve for RF win-loss model from (Ravari 2017).

Average precision is also a common performance measurement in machine learning where the order is important. Therefore, we used average precision to show how accurate the ranking model can predict the rank of a player in free-for-all matches.

Combined models

Table 2 shows the performance for the combined models. For the binary models, AUC and average precision tend to be close, but for multiclass models average precision is much lower than AUC. As Davis and Goadrich (2006) states, a precision-recall curve provides more insight into the accuracy for ranking problems. Thus, we compare the performances in terms of average precision, and later we discuss the precision-recall curves of our multiclass classification models.

As Table 2 shows, RF models outperform GB models in most classification tasks. Thus, we focus on RF models. The comparison of RF models show that the win-loss model, ranking model, and binary-ranking model achieved 84%, 68%, and 94% average precision respectively. It is not unexpected that a win-loss model outperforms a ranking model, as the ranking model has more classes.

Figure 1 precision-recall curve for RF win-loss model. As the plot shows, precision starts at 1 for recall 0, and steadily decreases to precision 0.5 at recall 1. In total, the model achieved 84% accuracy in terms of AUC.

Figure 2 show the precision-recall curve for the RF ranking model. In this Figure, classes 0, 1, and 5 have higher performances compared to the other classes. In the other words, the model predicts higher ranks (0 and 1) and the lowest rank (5) more accurately than the mid-ranks 2, 3, and 4.

Figure 3 summarizes the performance of the RF ranking models for each class by a normalized confusion matrix. The columns show predicted class labels, and rows represent the true class labels (Stehman 1997). For instance, the value in column 0 and row 1 is 0.05, which indicates that 5% of class 1 samples are assigned to class 0. As we saw before, ranks 0, 1, and 5 are predicted with high accuracy, while ranks 2, 3, and 4 are predicted with lower accuracy. Avontuur, Spronck, and Van Zaanen (Avontuur, Spronck, and Van Zaanen 2013)

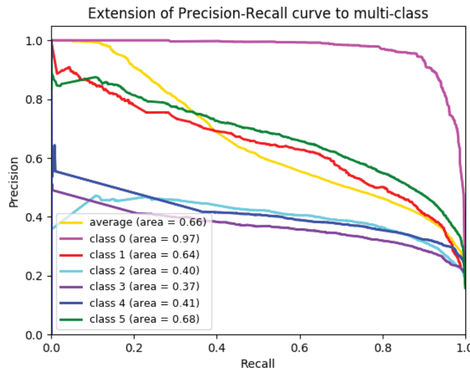

Figure 2: Precision-Recall curve in ranking model of RF classifier from (Ravari 2017).

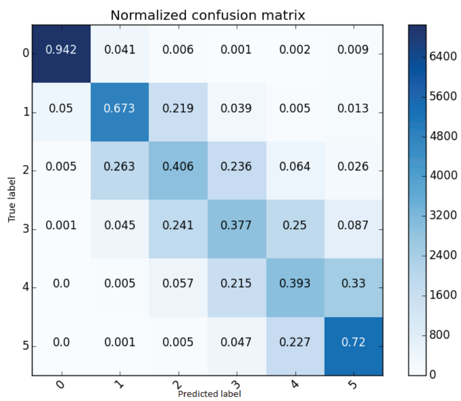

Figure 3: Normalized confusion matrix in ranking model of RF classifier from (Ravari 2017).

win-loss	ranking	binary-ranking
SPL (0.06)	K (0.12)	KD (0.6)
SPK (0.05)	SPL (0.11)	K (0.3)
D (0.04)	KD (0.10)	ADS (0.05)
KDA (0.03)	KDA (0.09)	KDA (0.04)
KD (0.03)	ADS (0.06)	TeamId (0.04)

Table 3: Top-5 features per RF models with their relative importance rate from (Ravari 2017). D:Deaths, K:Kills, KD: Kill death ratio, KDA: kills deaths assists, SPL: average score per life, ADS: activity duration seconds, SPK: average score per kill.

predicted league levels of *StarCraft* players with 44% accuracy on average, but with close to 90% accuracy for the best players. Similar to our ranking model, their model has higher performance in prediction of top and low levels, but it has lower performance in prediction of mid levels.

Table 3 summarizes the top-5 features and their importance rate for RF models. Feature importances were calculated according to Gini importance. The top-5 features in the win-loss model and ranking model are similar, but differently ordered. Score-Per-Life (SPL) has a high importance rate in both models. Kills (K) is the biggest difference in the list of top-5 features in these models. While Kills has the highest order in ranking model, it is not among the top-5 of the win-loss model. Indeed, in free-for-all game modes (which the ranking model encompasses) getting kills and avoiding being killed are of utmost importance. The third column of Table 3 shows top-5 features for the binary-ranking model. In comparison with the ranking model, Kill-Death Ratio (KD) is moved to the top of the binary-ranking model, SPL is missing, and Team Id is added to the top-5 features. A possible explanation for SPL being of lower importance for the binary-ranking model might be that it is mainly used to distinguish between ranks which are close together.

Game-mode	Classification	AUC	avg precision
Skirmish	GB-Binary	89%	89%
Control	GB-Binary	83%	82%
Salvage	GB-Binary	94%	94%
Clash	GB-Binary	77%	75%
Trials Of Osiris	GB-Binary	99%	99%
Doubles	GB-Binary	94%	94%
Iron Banner	GB-Binary	83%	82%
Elimination	GB-Binary	99%	99%
Rift	GB-Binary	80%	78%
Mayhem Clash	GB-Binary	75%	73%
Zone Control	GB-Binary	95%	94%
Rumble	GB-Multiclass	86%	84%
Supremacy	GB-Multiclass	98%	98%

Table 4: Performance of models per game mode from (Ravari 2017).

Individual models

The dataset includes 13 game modes. Supremacy and Rumble are ranking modes, the others are win-loss modes. Thus, we employed binary classifier for each win-loss mode, and multiclass classifiers for the ranking modes.

Table 4 shows the performance of individual models in terms of AUC and average precision. Since the performance of the RF classifiers was similar to the GB classifiers, we removed the RF models in this table. The comparison of individual win-loss models with the combined win-loss model in Table 2 in terms of average precision shows most of the individual models have higher performance except for the Clash, Rift, and Mayhem Clash models. The low performance of these models may be due to the fact that these three game modes represent 6v6 matches, while all the others represent 3v3 matches. Among the individual models, Trials Of Osiris, Elimination, and Supremacy models have a very high performance, around 99%. This may be explained by the fact that for these game modes the outcome is almost exclusively determined by the number of kills, which is one of the features in our dataset.

While most PvP modes are cooperative, our models were trained on the features of individual players. And yet the models can still predict match outcomes with high accuracy. This may be due to some features representing cooperative

Skirmish	Control	Salvage	Clash	Trials Of Osiris	Doubles	Iron Banner	Elimination	Rift	Mayhem Clash	Zone Control	Rumble	Supremacy
SPL(0.14)	SPL(0.15)	SPL(0.26)	SPL(0.10)	SPL(0.40)	SPL(0.17)	SPL(0.14)	SPL(0.44)	SPL(0.12)	SPL(0.08)	SPL(0.40)	K(0.12)	K(0.39)
D(0.14)	SPK(0.13)	D(0.10)	D(0.08)	SPK(0.19)	D(0.15)	SPK(0.13)	SPK(0.17)	KDA(0.09)	D(0.08)	SPK(0.10)	KDA(0.10)	SPL(0.12)
KDA(0.90)	KDA(0.07)	SPK(0.09)	KDA(0.07)	D(0.17)	KDA(0.13)	KD(0.07)	D(0.10)	D(0.07)	KDA(0.08)	D(0.07)	SPL(0.08)	KD(0.09)
KD(0.06)	ADS(0.06)	KD(0.07)	ADS(0.06)	KD(0.07)	KD(0.09)	KDA(0.07)	KDA(0.06)	SPK(0.07)	SPK(0.07)	ADS(0.05)	SPK(0.08)	SPK(0.09)
ADS(0.06)	D(0.06)	KDA(0.06)	SPK(0.06)	KDA(0.04)	ADS(0.06)	ADS(0.06)	KD(0.05)	ADS(0.07)	ADS(0.06)	KD(0.05)	KD(0.07)	D(0.07)

Table 5: Top-5 features per RF models with their relative importance rate from (Ravari 2017). D:Deaths, K:Kills, KD: Kill death ratio, KDA: kills deaths assists, SPL: average score per life, ADS: activity duration seconds, SPK: average score per kill.

performance metrics, such as SPL and KDA. SPL is the sum of scores that a player earned during his life that includes co-operative actions such as assist, revive, and capture a zone. KDA also includes assists. As Table 5 shows, SPL and KDA have an important role in all PvP game modes.

Table 5 shows top-5 features for individual models. In all of the win-loss models, SPL is the strongest predictive feature, while in ranking models kills is the strongest predictive feature. SPK and deaths are also strong predictive features in win-loss models. Top-5 features in win-loss models are very similar, with different orderings. For the two ranking models, the list of top-5 features is also quite similar. The weight of SPL is especially high for Elimination (0.44), Trials Of Osiris (0.40), and Zone Control (0.40). In these game modes, players must capture a zone or kill all of the opponents. KDA is the most frequent feature in top-5 features after SPL. KDA is found in the most of the individual models, except Zone Control and Supremacy models. In Skirmish, Salvage, Clash, Doubles, and Mayhem Clash models, deaths is the second strongest predictive feature. In these game modes, keeping teammates alive is critical. In general, kills, deaths, KD, KDA, SPL, ADS, and SPK are the most important player's performance metrics in different PvP game modes. Kills, deaths, and KD show how much a player is involved in fighting other players. KDA also reflects cooperation between team members in addition to kills and deaths. SPL represents how much the player earned points during his life. Players can earn points from activities other than kills, deaths, and assists, namely actions such as capturing, neutralizing, or defending a zone, and reviving a teammate. Thus, SPL includes scores that are related to cooperation. ADS shows how long players spend time in a match. SPK shows the points that a player gets for kills. A high value may entail that a player often manages to pull off complex kills such as headshots, or kills using melee weapons or grenades. Most of the player's performance metrics that evaluated in this study are available in the other combined MMO games and FPS games. We expected that a similar approach would work for these games.

To sum up, the results shows that match result prediction is possible in *Destiny*. In win-loss matches, the models predict the winner with an accuracy higher than 80%. In ranking matches, where six outcome classes exist, the models' prediction accuracy is at least 68% in terms of average precision. In ranking matches, top and bottom ranks can be predicted with higher accuracy than the mid-ranks. As expected, the individual models have higher performance compared to the combined models. Interestingly, some of the individual models predicted the match results by 99%, i.e., almost perfectly. The comparison of top performance metrics in win-loss models and ranking models shows that in ranking game modes kills is the most important player performance metric to get the best result, while in win-loss modes avoiding to die is more important. In individual models, the top-5 performance metrics are almost the same, but with different orderings in different game modes. SPL is the strongest predictive feature in both win-loss matches and in ranking matches. KDA is the second strongest predictive feature across game modes. Both of these metrics integrate elements of cooperation. Generally, players seem to focus on the actions that earn more points in different game modes (which comes at no surprise).

Discussion and Conclusion

In this paper prediction models are presented for the major commercial console title *Destiny*. Based on a dataset of match records and other behavioral metrics from *Destiny*, we developed combined classification models which encompass all PvP game modes, including both win-loss matches and ranking matches. The models achieve performance between 68% and 99%. The results suggest that match prediction in competitive multi-player shooter games (which includes major e-sports titles such as *CounterStrike* and *Team Fortress 2*), should be performed on specific game modes, as the models developed for the individual PvP game modes of *Destiny* outperform combined models. This results also highlights that players adapt their behavior to the conditions of the different games modes. The comparison of top-5 player performance features between models build for win-loss game modes and ranking game modes, i.e., combining data from all PvP modes in the game, shows that SPL, KDA and KD are the most important player performance metrics. In ranking models, player kills (K) is the most important. Furthermore, we compared player performances across PvP game modes by individual models, and found that SPL and KDA are the most important features. As we examined a variety of performance metrics across a set of PvP game modes, it might be possible to extrapolate results to other FPS games - and potentially be employed to customize game modes to challenge different performance vectors. For future research, we are interested in predicting the match results across time to analyze players' progress, and relate patterns in progress to match results and player performance. In the present work, we studied post-match features, but there is an opportunity to integrate within-match features which potentially could be used to increase prediction accuracy. It might be possible to combine action sequence or encounter derived features with post-match features in match prediction.

Acknowledgements

Part of this work was conducted in the Digital Creativity Labs (www.digitalcreativity.ac.uk), jointly funded by EPSRC/AHRC/InnovateUK under grant no EP/M023265/1.

References

Avontuur, T.; Spronck, P.; and Van Zaanen, M. 2013. Player skill modeling in starcraft ii. In *Artificial Intelligence in Interactive Digital Entertainment Conference*.

Bakkes, S.; Spronck, P.; and van den Herik, J. 2007. Phase-dependent evaluation in rts games. In *19th Belgian-Dutch Conference on Artificial Intelligence*, 3–10.

Breiman, L. 2001. Random forests. *Machine learning* 45(1):5–32.

Cole, N.; Louis, S. J.; and Miles, C. 2004. Using a genetic algorithm to tune first-person shooter bots. In *Evolutionary Computation Congress*, volume 1, 139–145. IEEE.

Davis, J., and Goadrich, M. 2006. The relationship between precision-recall and roc curves. In *Proceedings of the 23rd international conference on Machine learning*, 233–240. ACM.

Drachen, A.; Lundquist, E.; Kung, Y.; Rao, P.; Sifa, R.; Julian, R.; and Klabjan, D. 2016. Rapid Prediction of Player Retention in Free-to-Play Mobile Games. In *Proc. of AAAI AIIDE*.

Erickson, G. K. S., and Buro, M. 2014. Global state evaluation in starcraft. In *Artificial Intelligence in Interactive Digital Entertainment Conference*.

Fogarty, J.; Baker, R. S.; and Hudson, S. E. 2005. Case studies in the use of roc curve analysis for sensor-based estimates in human computer interaction. In *Proceedings of Graphics Interface 2005*, 129–136. Canadian Human-Computer Communications Society.

Friedman, J.; Hastie, T.; and Tibshirani, R. 2001. *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin.

Friedman, J. H. 2002. Stochastic gradient boosting. *Computational Statistics & Data Analysis* 38(4):367–378.

Hsu, C.-J.; Hung, S.-S.; and Tsay, J.-J. 2013. An efficient framework for winning prediction in real-time strategy game competitions. *Innovation, Communication and Engineering* 239.

Ravari, Y. N.; Bakkes, S.; and Spronck, P. 2016. Starcraft winner prediction. In *Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Ravari, Y. N. 2017. http://gameanalysis.wikidot.com/plots. licensed under CCA 4.0 International license.

Schatten, M.; Tomičić, I.; and Durić, B. O. 2015. Multi-agent modeling methods for massivley multi-player on-line role-playing games. In *International Information and Communication Technology, Electronics and Microelectronics Convention on*, 1256–1261. IEEE.

Schubert, M.; Drachen, A.; and Mahlmann, T. 2016. Esports analytics through encounter detection.

Sifa, R.; Hadiji, F.; Runge, J.; Drachen, A.; Kersting, K.; and Bauckhage, C. 2015. Predicting Purchase Decisions in Mobile Free-to-Play Games. In *Proc. AAAI AIIDE*.

Sifa, R.; Srikanth, S.; Drachen, A.; Ojeda, C.; and Bauckhage, C. 2016. Predicting retention in sandbox games with tensor factorization-based representation learning. *Computational Intelligence in Games*.

Stehman, S. V. 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote sensing of Environment* 62(1):77–89.

Yang, P.; Harrison, B. E.; and Roberts, D. L. 2014. Identifying patterns in combat that are predictive of success in moba games. In *Foundational of Digital Games Conference*.