



Determinants of victory in Esports - StarCraft II

Andrzej Białecki¹ · Jan Gajewski² · Piotr Białecki³ · Ashwin Phatak⁴ · Daniel Memmert⁴

Received: 27 May 2021 / Revised: 8 February 2022 / Accepted: 15 May 2022 /
Published online: 17 September 2022
© The Author(s) 2022

Abstract

Esports offer a unique opportunity to conduct human performance studies, as they use modern hardware and software as an operation platform. Insights on gameplay and underlying processes may push the development of new and optimal practice methods. The aim of this study was to investigate performance indicators from in-game data to predict the outcome of the matches in StarCraft II: Legacy of The Void.

Data from 6509 games (game records) provided by 5 players at the level of Master or GrandMaster were used. The distribution of analyzed players concerning the preferred in-game race was as follows: “Protoss” (n=3), “Zerg” (n=1), “Terran” (n=1). Each game record contained data for both the winner and the loser. In total, 3719 game records and 9 performance indicators were obtained after applying the inclusion criteria.

Logistic regression with 5-fold cross-validation was performed to predict the game outcome. The model was able to discriminate the game outcome (won, lost) with an out-of-sample accuracy of 0.728 ± 0.021 . The performance indicators which showed the strongest effect in predicting the game outcome were “minerals lost army” [p-value<0.001, std_odds_ratio: 0.069], “minerals killed army” [p-value<0.001, std_odds_ratio: 6.446], “minerals used current army” [p-value<0.001, std_odds_ratio: 4.081], and “minerals killed economy” [p-value<0.001, std_odds_ratio: 2.896]. It seems evident that winner optimized interaction with an opponent by keeping his/her own army intact while inflicting damage to the opponent’s army or economy. In conclusion, the effective use of the army, based on optimizing the ratio between units lost and units killed, may be significant in predicting the game outcome.

Keywords Esports · Performance analysis · Gaming · Starcraft · Real-time strategy

1 Introduction

Esports are modern sports carried out on different technical platforms at the forefront of measuring performance in human-computer interaction. The competitive nature of esports

✉ Ashwin Phatak
a.phatak@dshs-koeln.de

may help with testing new methods of measurement. Esports inherently use the same computer input devices as in different human-computer interaction studies [15]. The international presence of esports and their popularity in the last decade can be attributed to technological development in hardware and software. Introduction of technology and new methods of activating users allowed for live broadcasting of gameplay and increased viewership outside traditional channels, which caused a move away from passive viewing of games to a more interactive environment due to the rise of streaming platforms [11, 21]. This development has led to academic interest in esports. So far, esports have been studied from a theoretical standpoint with the primary argument being in the classification of esports as “real sport” [24].

Sports analytics is widely recognized and used in different contexts. Studies indicate that correct use of available data may provide a competitive advantage and better game understanding [40]. Due to the requirement of a technological platform, nearly all electronic sports have certain features that allow for a thorough statistical analysis of the game. Games (software) are simulated environments. Such games save the data allowing a game to be recreated using provided tools, features, or game engines. Due to the virtual nature of esports, it is possible to obtain highly specific data at high volume and resolution. This provides a unique opportunity to explore resulting “Big Data” using not only traditional statistics but also artificial intelligence (AI) and machine learning (ML) tools, thus uncovering its potential to understand the performance in esports and provide insights on human-computer interaction (HCI) [5].

1.1 History of Esports

Due to the limited availability of games and technology, esports in its simplest form began at universities [44]. The first esports competition dates back to the 1970s. The first esports tournament was called the Intergalactic Spacewar Olympics. It was played at Stanford University in the United States on October 19, 1972 [17]. Some sources on the history of electronic sports do not mention the games held in 1972 and consider The Space Invaders Championship - a tournament with an estimated 10,000 spectators - to be the first. The aforementioned tournament was played using the title Space Invaders [12, 16]. Information on the beginnings of electronic sport and its presence in the literature is quite limited. The first “electronic sportsmen” were certainly scientists, university faculty employees and students [9].

1.2 Public perception of Esports

The market value data shows a constant increase and estimated further growth and development of the esports industry [8]. Surveys indicated more than one million live broadcasts on Twitch in the period under review from September 29th, 2011 to January 9th, 2012. Figure 1 shows higher average viewership coinciding with the broadcasting of esports events.

The statistics of the World Cyber Games in the years 2000 to 2005 show incremental growth in terms of the number of people participating in the games, the number of countries represented, and the associated prize pool, as displayed in Table 1 [20]. Data on prize money in electronic sports were also collected. A perfect example of such a project is “e-Sports Earnings” - the database contains statistics dating from 1998. Statistics from 2017 indicate more than 113 million U.S. dollars in prizes, although the credibility of this source is unknown [14]. On the other hand, it is possible to verify and assess the total prize pool of major esports events such as The International (Dota 2), which shows constant growth of

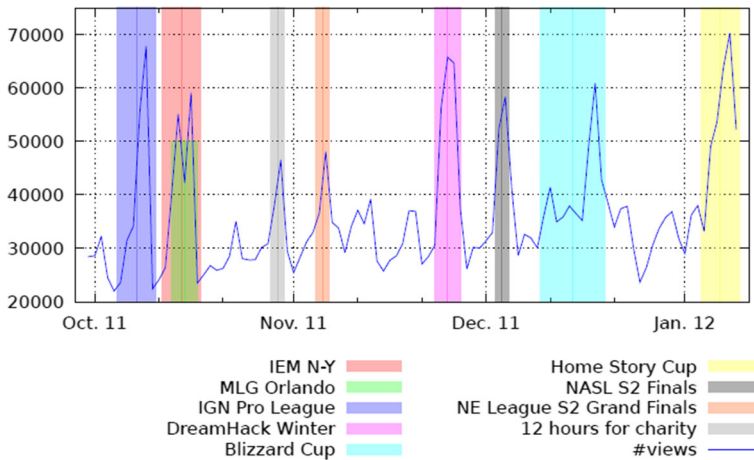


Fig. 1 Average number of viewers in the examined period together with the most important sporting events [25]

prize pools in the years 2012 to 2021 approximately between 1.6 and 40 million USD. This can be viewed in contrast to the historical data displayed in Table 1 [31].

Sources indicate that the audience of electronic sports was influenced by the fact that the recipients were often personally involved in the competition at a level lower than or equal to the one presented during the tournaments in esports, thus fulfilling the aspect of socialization in the players’ environment [27].

Despite all the evidence suggesting increasing popularity and professionalism in esports, the academic research on the topic is primarily theoretical. Multiple studies argue whether esports can be defined as a “sport” [19, 24, 45].

1.3 Research on gaming performance

Study on exploration of virtual game environments examined the differences in motivation by defining four different archetypes based on their exploration approach [42]. Sources on game analysis have addressed the subject of data collection and processing in the context of AI. Such research was conducted with the goal of improving AI based opponents or evaluating their performance [3, 26]. Publications analyzing how games were played in electronic sports attempted to find factors that may influence victory or defeat, thus describing the most desirable course of competition for an electronic sportsman. A study

Table 1 Tournament data for World Cyber Games 2000-2005 [20]

Year	2000	2001	2002	2003	2004	2005
Location	Seoul, Korea	Seoul, Korea	Daejeon, Korea	Seoul, Korea	San Francisco, United States	Singapore
No. of Countries Represented	17	37	45	55	59	70
WCG Participants	174	389	456	562	642	Approx. 800
Total Prize Money (Approx.), in USD	200 000	300 000	300 000	350 000	420 000	435 000

involving the game Dota 2 emphasized the predictors of victory. The researchers analyzed interactions between rival teams through game record data analysis [41]. Publications on the collection and use of data from games mention maximizing the chances of winning after changes resulting from data analysis [7]. Gameplay research conducted in the first expansion of the StarCraft game identified problems related to the prediction of the winner using AI tools. Artificial intelligence achieved similar results of prediction as human “experts”. It was found that the humans were considering a different set of variables that could not be analyzed by AI in this specific study. Furthermore, methods used were related to the creation of “influence maps” assigned to the units present within the game, determined through data analysis [39]. Authors researching SC2 point to the value that available data can bring and their application in game modeling and description. Simulating datasets and training classifiers is a viable strategy for research of simple combat outcomes [28]. The research conducted made it possible to find trends related to the application of specific tactics by analyzing player behavior based on a sequence of created structures. It is worth noting that the authors also verified application of specific sequences and their relationship with victory. However, the methodology used excluded basic variables present in the game [33]. Due to the constant progress in esports and associated technology, the need for future research in esports and modeling game characteristics can be noted [23]. Part of the research was conducted using older versions of the game (StarCraft, StarCraft: Brood War, StarCraft II: Wings of Liberty, StarCraft II: Heart of the Swarm). Such studies could be reproduced in the newest version of the game: StarCraft II: Legacy of The Void, which was released in 2015 with major gameplay changes [13, 26, 33, 39, 47].

Sources on computer games dealt with data collection and processing. However, so far, they have not been able to solve the problem comprehensively in the field of esports to provide easily applicable insights. There have been few studies that analyzed player performance in different games. AI has been used to improve the game design and compare outcome prediction between human “experts” and AI based tools. There seems to be limited research describing key performance indicators in elite esports athletes in StarCraft II: Legacy of the Void. The research gap in analyzing performance of esports athletes is noticeable [38]. The current study intends to address the research gap from a practical perspective of human-computer interaction and sport science. Based on the point of view of athletes, it seems to be appropriate to start the research by verifying the key aspects of the game with computational methods.

The present study aims to investigate performance indicators which significantly differentiate winners and losers in the game of SC2. The following questions were prepared as the basis for further research:

- **RQ1** Is logistic regression a sufficient method to model the game-engine generated data to differentiate between player distributions based on the game outcome?
- **RQ2** Which of the variables provide a computationally observable distinction between winners and losers in StarCraft II: Legacy of The Void?

2 Methods

2.1 Inspected game characteristics

StarCraft II: Legacy of The Void (SC2) contains various game modes: 1v1, 2v2, 3v3, 4v4, Archon, and Arcade. The most competitive and esports related mode (1v1) can be classified

as a two-person combat, real-time strategy (RTS) game [42, 43]. The goal of the game for each of the competitors is either to destroy all of the structures, or to make the opponent resign.

The outcome of a single SC2 game can end in three ways: win, loss, or draw. Winning or losing a game is achieved either by all of the player's structures getting destroyed, or a surrender. A draw occurs when none of the players perform the actions recognized by the stalemate system built into the game for 129 seconds [30]. Before starting the game, players decide the "race" they will be playing. Possible choices are: "Terran", "Zerg", and "Protoss".

The game is initiated by starting the search for an opponent, with the following possible variants: Ranked game - Players use a built-in system that selects their opponent based on Matchmaking Rating (MMR) points [29]. Unranked game - Players use a built-in system that selects their opponent based on a hidden MMR - such games do not affect the position in the official ranking. Custom game - Players join the lobby (game room), where all game settings are set and the readiness to play is verified by both players - this mode is used in tournament games. Immediately after the start of the game, players have access to one main structure, which allows for further development and production of units. Figure 2 presents the appearance of basic structures for each race.

Each of the races has different gameplay mechanics. Moreover, the races differ in technological development. The differences in gameplay can be seen at the beginning of the game. The basic Protoss unit - Probe - can warp in structures (buildings), which do not require the presence of any unit to finish constructing. The only condition for the creation of a building, in this case, is its initialization in a specific area designated by a Pylon, which provides power to the structures of this race. If the Pylon is destroyed, all structures without power are unable to sustain production. This condition is true for all static structures except the main building of this race, i.e. Nexus, and for the structure in question - Pylon. Moreover, all races in the game have a maximum unit limit. Traditionally, during the 1 vs 1 competition increase of the unit limit is achieved by building a structure, or a specialized unit. In the case of Protoss, it is a Pylon. The basic Zerg unit is called a Drone and is designed to create structures. Completion of the structure results in the loss of the basic unit, which turns (evolves) directly into a building. Buildings of this race can only be constructed on creep. Creep is formed by the basic structure of this race, i.e. Hatchery, or utilizing special units and structures (Queen, Creep Tumor). A different aspect of Zerg is that to enlarge the unit



Fig. 2 Basic structures - "bases" for each race available in StarCraft II in the following order (from the left): "Terran", "Protoss", "Zerg"

limit of Zerg one must create a specialized unit - Overlord. The basic Terran unit, SCV, initiates the construction of a structure and stays with it until the construction is completed. Exceptions are cases when a player gives an order to stop the construction of a building or an order to cancel the construction. The basic structure of the Terran is the Command Center. Some structures of the described race can break away from the ground. Increasing the limit of units requires a structure - a Supply Depot. A detailed description of all units available in StarCraft II can be found on the community-managed website of Liquipedia [32].

2.2 Sample

Data from 6509 games obtained from 5 players at the level of Master or GrandMaster were used with consent for scientific publication. Respondents belonged to approximately the best 5% of players in the world. Global player ranking statistics are available on platforms such as RankedFTW [37] or for tournament games Alligulac [2]. Each game record contained data for both the winner and the loser, thus substantially expanding the sample size given that the opponents are selected based on MMR points and are likely to be different players of similar expertise with their own preferred races which were not verified. The distribution of analyzed players concerning the preferred in-game race was as follows: “Protoss” (n=3), “Zerg” (n=1), “Terran” (n=1). Opponents of the examined players were selected according to the ranking games system based on MMR points.

Extracting data was performed using Python 2.7 with an open-source library provided by the game developers - Blizzard Entertainment [6]. Initial extraction yielded data from 5937 games played in 1v1 mode.

2.3 Data Pre-processing

Definitions for all of the inspected variables are available in Appendix A. The following inclusion criteria were applied for filtering the relevant information:

- ranked play (variable competitive equal to 1) - Data were a part of competitive environment that influenced players’ rating.
- games lasting more than 90 seconds - Games lasted long enough for the players to interact.
- variable APM higher than 0 - Players must have performed more than 0 actions.
- variable MMR higher than or equal to 0 - Ensured the deletion of data artifacts.
- all other variables higher than or equal to 0 - All of the performance indicators selected for inspection were positive.

Player race and inclusion criteria filtering information is presented in Table 2.

Any replay that did not meet specified criteria was removed from the dataset. After applying the inclusion criteria, data from 3719 games were further processed.

The analyses were carried out using game state information. All of the variables which were available for the analysis are listed in Appendix B. A total of 9 out of 41 variables were selected. Exclusion of multicollinear performance indicators was performed. Game records contained pre-computed data saved every 7 seconds in the form of the game state that occurred at the time of saving. The game state data present were expressed in a unit of time “gameloop” (See Appendix A). In addition, pre-computed values for APM and MMR were available for each game. Replays under study were characterized by a high standard deviation (SD) of the game’s duration. It was found that the minimum value of the game

Table 2 Race distribution of games and filtering information for inclusion criteria

Player	Before Filter	After Each Filter					Cumulative
	Games	Games					Games
		ranked play	time	APM	MMR	other	
Protoss (n = 3)							
Player 1	1608	1498	1473	1473	1451	1122	1122
Player 2	1918	1414	1398	1398	1379	1132	2254
Player 3	852	590	585	585	581	437	2691
Zerg (n = 1)							
Player 4	1162	1162	1145	1145	1115	829	3520
Terran (n = 1)							
Player 5	397	317	305	305	303	199	3719

duration was at the limit of the criteria. On average games lasted 744.96 ± 291.31 s. Selected indices after exclusion are shown in Table 3.

2.4 Statistical analysis

Logistic regression was used to model for the result of the game (won, lost). Balancing was not required as the data for both examined finishing game states were recorded symmetrically. Testing out-of-sample validity of current findings was performed by using 5-fold cross-validation. Pre-processing and statistical analysis were performed using Python 3.7.7 (see the code).

3 Results

3.1 Pre-analysis

To better understand the difference between the final game conditions (won, lost), pre-analysis was conducted. In Table 3 average results (\pm SD) are presented with ranges of values for variables used in the logistic regression modeling with the exception of “gameloop”, which does not differentiate between winning and losing. The variable “minerals used current army” differentiated between victory and defeat by an average of 271.18 minerals. It was found that the winning players lost on average 137.66 minerals per army unit less than the players who were defeated.

3.2 Logistic regression

For further modelling and inspection, logistic regression was used. Results of logistic regression conducted on the chosen performance indicators are displayed in Table 4. It was found that one of the selected variables (APM) did not significantly differentiate between the analyzed game outcomes (won, lost) with a p-value >0.05 . Every other variable was found to significantly differentiate between winners and losers at different thresholds of p-value <0.05 , <0.01 , <0.001 . The analysis of selected performance indicators confirmed that the

Table 3 Descriptive statistics for analyzed data (n = 3719)

Analyzed Parameter	Average		Min		Max	
	won	lost	won	lost	won	lost
game time [s]	744.96 ±291.31	744.96 ±291.31	169.64	169.64	3857.41	3857.41
minerals used current army	1546.46 ±830.48	1275.28 ±769.90	19.57	20.45	7199.10	8564.11
minerals collection rate	1620.59 ±378.34	1482.60 ±358.96	323.22	229.66	2671.51	2487.57
workers active count	41.07 ±10.90	37.75 ±10.24	6.44	5.86	70.46	65.66
food made	92.40 ±34.64	85.54 ±32.83	19.94	20.69	282.45	255.74
food used	71.08 ±27.10	62.95 ±24.16	12.15	11.34	166.66	143.81
minerals killed economy	521.21 ±504.12	356.78 ±446.51	1.11	0.00	5353.46	7107.64
minerals lost army	1614.34 ±1920.67	1752.00 ±1953.26	4.24	4.44	25742.80	32631.19
minerals killed army	1756.52 ±1955.29	1619.25 ±1926.27	4.44	4.24	32631.19	25742.80
MMR	5926.77 ±351.47	5790.31 ±350.58	4007.00	3598.00	7202.00	6903.00
APM	255.81 ±65.68	246.58 ±61.33	100.00	68.00	709.00	530.00
minerals current	368.71 ±290.89	321.79 ±197.91	63.52	48.14	8769.99	2787.64

highest standardized coefficient magnitude and standardized odds ratio with respect to victory were observed for “minerals lost army” [std_coefficient: -2.675, std_odds_ratio: 0.069]. In contrast, the variables “minerals killed army” [std_coefficient: 1.864, std_odds_ratio: 6.446] and “minerals used current army” [std_coefficient: 1.406, std_odds_ratio: 4.081] were found to be positively associated with victory. The rest of the observed selected variables concerned parts of the game economy with standardized coefficient magnitudes regarding victory ranging from -2.146 to 1.610.

The logistic regression model showed high accuracy, AUC, and AUCPR. Furthermore, the low standard deviation in cross-validation results shows that the model was stable and reliable for out-of-sample validity. Selected performance metrics (\pm SD) are presented in Table 5.

In Fig. 3 the receiver operating characteristic (ROC) curve is displayed for the model with the AUC metric obtained on the testing set. The plot shows the diagnostic ability of the trained logistic regression model at different classification thresholds.

Table 4 Logistic regression coefficients and distribution metrics

Name	coefficient	std_error	z_value	p_value	std_coefficient	odds_ratio (for std_coefficient)
Intercept	-7.52105	0.68655	-10.955	0.000000e+00	0.033	1.034
APM	-0.00002	0.00071	-0.031	9.756562e-01	-0.001	0.999
MMR	0.00087	0.00012	7.251	4.136691e-13	0.309	1.362
Average minerals collection rate	0.00421	0.00039	10.764	0.000000e+00	1.610	5.003
Average minerals current	0.00040	0.00018	2.173	2.974358e-02	0.110	1.116
Average minerals lost army	-0.00128	0.00008	-16.287	0.000000e+00	-2.675	0.069
Average minerals killed army	0.00091	0.00007	12.527	0.000000e+00	1.864	6.446
Average food made	-0.06231	0.00578	-10.777	0.000000e+00	-2.146	0.117
Average food used	0.01010	0.00357	2.829	4.675667e-03	0.265	1.304
Average minerals used current army	0.00169	0.00016	10.587	8.264943e-10	1.406	4.081
Average workers active count	-0.05182	0.01641	-3.157	1.591768e-03	-0.562	0.570
Average minerals killed economy	0.00211	0.00014	15.605	0.000000e+00	1.063	2.896

Table 5 Train, test and cross-validation metrics

Metric	Train	Cross-Validation±SD
Accuracy	0.742	0.728±0.021
AUC	0.800	0.798±0.021
AUCPR	0.798	0.798±0.037

Confusion matrices obtained in training and cross-validation, and used to calculate the metrics of the model, are shown in Table 6. Information Displayed is the number of true negatives, false positives, false negatives, and true positives classified for possible outcomes (won, lost) predicted in the process of training and cross-validation.

4 Summary

The Availability of modern analytical tools may increase player performance if used correctly [1, 18, 35]. The use of new technologies is at the heart of electronic sports. Using game engines to simulate a game environment creates the possibility to directly interfere with data available immediately after the game is over [10, 36]. In some cases, it might be possible to interface directly with the game server or the game engine to provide real-time information on player performance, as shown by the research done on reinforcement learning [22]. Video games constituting a platform for esports differ from other software as a medium. It seems that esports are a perfect bridge between sports science and computer science in the specific areas of human-computer interaction and sports analytics, among others [4].

We claim that the results obtained by using logistic regression were sufficient to describe the key performance indicators (determinants of victory) in StarCraft II: Legacy of the Void. This statement provides answer to the research question **RQ1**: “Is Logistic Regression a sufficient method to model the game-engine generated data to differentiate between player distributions based on the game outcome?”. Moreover, all of the inspected variables provide meaningful distinction between winning and losing players in our dataset, simultaneously

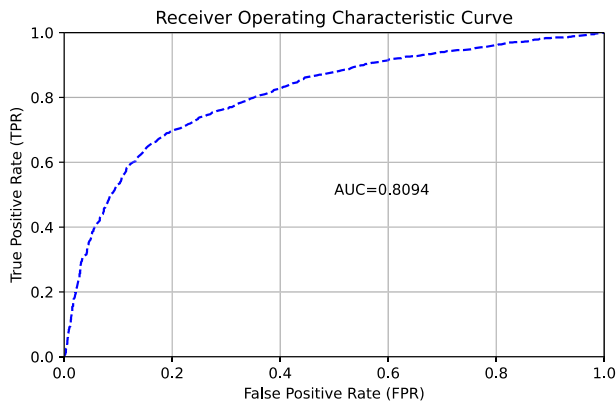


Fig. 3 ROC curve generated by H2O in the process of testing. Area under the ROC curve (AUC) = 0.8094

Table 6 Confusion matrices for training at max f1 @ threshold = 0.457, testing at max f1 @ threshold = 0.384, and cross-validation at max f1 @ threshold = 0.467

	lost	won	Error	Rate
Train				
lost	1274.0	580.0	0.313	(580.0/1854.0)
won	419.0	1459.0	0.223	(419.0/1878.0)
Total	1693.0	2039.0	0.268	(999.0/3732.0)
Test				
lost	1034.0	831.0	0.446	(831.0/1865.0)
won	255.0	1586.0	0.139	(255.0/1841.0)
Total	1289.0	2417.0	0.293	(1086.0/3706.0)
Cross-Validation				
lost	1306.0	548.0	0.296	(548.0/1854.0)
won	445.0	1433.0	0.237	(445.0/1878.0)
Total	1751.0	1981.0	0.266	(993.0/3732.0)

answering the research question **RQ2** that asks: “Which of the variables provide a computationally observable distinction between winners and losers in StarCraft II: Legacy of The Void?”. This is described in detail in Section 4.1

4.1 Key findings

The present study aimed to analyze selected performance indicators (see Appendix A) that allow for player differentiation concerning the game outcome. Key findings were split into three categories based on in-game mechanics. The first category, “micro”, is the results that concern army units and their utilization. The second category, “macro”, is the results that concern the in-game economy. The third category is the overall performance indicators. Key findings are associated with standardized coefficient, and standardized odds ratio values. Standardized coefficients explains the change in dependent variable - game outcome prediction (won, lost) for increase of one standard deviation change in measured independent variable (performance indicator) distribution. Standardized odds ratio explain the change in odds of classifying the game outcome (won, lost) for increase of one standard deviation change in measured independent variable (performance indicator) distribution [34].

All of the variables from the first category were found to be statistically significant with p-values <0.001 for “minerals lost army” [std_coefficient: -2.675, std_odds_ratio: 0.069], “minerals killed army” [std_coefficient: 1.864, std_odds_ratio: 6.446], “minerals used current army” [std_coefficient: 1.406, std_odds_ratio: 4.081], and “minerals killed economy” [std_coefficient: 1.063, std_odds_ratio: 2.896]. It seems evident that the winner optimized interaction with an opponent by keeping his/her own army intact while inflicting damage to the opponent’s army or economy. These results highlight the importance of proper unit control at the highest level of play.

The second category was found to have some variables with p-values suggesting lower significance with thresholds of p-value <0.05, <0.01. Standardized coefficient values were “minerals collection rate” [std_coefficient: 1.610, std_odds_ratio: 5.003], “food used”

[std_coefficient: 0.265, std_odds_ratio: 1.304], “minerals current” [std_coefficient: 0.110, std_odds_ratio: 1.116], “workers active count” [std_coefficient: -0.562, std_odds_ratio: 0.570], and “food made” [std_coefficient: -2.146, std_odds_ratio: 0.117]. The results suggest that the most important part of the in-game economy is obtaining minerals that can be spent either to further expand the economic potential, or to build more army units. Another important highlight is that the variable “workers active count” specifying how much mineral harvesting units were created seems to negatively impact the chance of winning. This suggests that optimizing the number of harvesting units versus the limit of 2 workers per mineral patch is important and that any disruptive action in the area of harvesting minerals can be used to gain an advantage.

In the third category containing overall performance indicators one of the variables to be insignificant i.e. APM (p-value >0.05), meaning that both winner and loser had similar ability in performing actions. This suggests that at the analyzed level of play the quality of actions is more important than their volume. The second variable in the category of overall performance indicators i.e. MMR (p-value <0.001), differentiated players with the following values: std_coefficient: 0.309, std_odds_ratio: 1.362. This result suggests that even though analyzed players were at the highest level of play, referred to as Master and Grand-Master, the internal system selected players with varied experience levels attributed to the same broader category.

A Different study highlighted other performance indicators which differentiate players through broad skill categories present in the in-game system: Bronze, Silver, Gold, Platinum, Diamond, Master, and a category not available in-game - professional. Currently, the system offers one more category of expertise - GrandMaster - which is associated with the top 200 players on a given server. Their study suggests that the value of APM is ranked as the first performance indicator that differentiates between levels of Bronze players and professionals. On the other hand, the value of “workers active count”, which is close to their performance indicator “workers made”, showed the biggest difference between Bronze and Gold levels of play [46].

4.2 Limitations and future work

The authors highlight the most basic, yet highly descriptive data. This made it possible to provide basic insights towards key aspects of the game using available information. Most of the in-game decisions derive from data that were analyzed. The present study shows different key performance indicators in the game of StarCraft II, and how they might be explored, but it is not without limitations. The remaining variables present in StarCraft II, which were not analyzed, should be considered to expand knowledge on more nuanced information, such as game tactics and precise unit control. Such nuanced information should also include gameplay differences based on server, maps, and preferred race. Ultimately real-time game-engine generated data provide the ability to perform sequence analysis on unit control and decision-making sequences. Such research could uncover time-related insights and pivotal points of the game.

It is recommended to undertake further research to find more factors and variables that explain different key components of gameplay in esports. In our study, we have only analyzed StarCraft II: Legacy of the Void. It is recommended to compare our results by researching other games. Additionally, we recognize that logistic regression is one of many classification methods, we recommend using other models and datasets to recreate, compare

and verify what was described in our work. Moreover, future research should analyze players at different levels of gameplay to develop optimal training methods aimed at maximizing the individual potential and, ultimately, competition outcome.

5 Conclusions

Based on the evidence presented, the following conclusions were drawn:

- The method of logistic regression allows one to find the key performance indicators (determinants of victory).
- It is possible to find the determinants of victory in esports using computational methods and the tools available in statistics, provided that the available data are properly processed beforehand.
- The main determinant of victory was found to be “minerals killed army”.
- The Average value of all the positive performance indicators was higher for winners.

In conclusion, the characteristics of players in terms of winning the competition at the highest level can be used as a model to follow.

Appendix A: Definitions

- APM - “Actions Per Minute” is a variable present in the game record, meaning the average number of actions performed by players in each minute of the game.
- MMR - “Matchmaking Rating” is a variable that indicates the number of ranking points of a given player. According to this variable, players are selected for competing within the internal game system.
- total_gameloop - the total game duration, in gameloop units. One second is equal to 22.4 gameloop units.
- gameloop - unit of time as recorded by the game engine.
- minerals_collection_rate - the average observed state of mineral extraction per minute. Minerals are the basic currency used to construct structures.
- minerals_current - a record of the current state of the minerals.
- minerals_lost_army - a record expressing the loss of minerals that were used to construct an army unit. It occurs when a player’s army unit has been destroyed.
- minerals_killed_army - a record expressing the value of enemy army units (in minerals) that a player has destroyed.
- food_made - the limit of units that can be created by a given player.
- food_used - record expressing the used limit of units of a given player.
- minerals_used_current_army - a variable expressing the sum of minerals used for units of the player’s army.
- workers_active_count - number of basic mineral extraction units.
- minerals_killed_economy - is a record expressing the value (in minerals) of enemy mining units and bases destroyed by the player.

Appendix B: Available variables

All of the available game state variables: total_gameloop, gameloop, minerals_used_current_technology, minerals_lost_technology, vespene_used_current_technology, vespene_current, vespene_used_in_progress_technology, minerals_collection_rate, minerals_current, minerals_lost_army, minerals_used_current_economy, vespene_friendly_fire_economy, vespene_used_current_economy, minerals_used_in_progress_army, minerals_killed_army, vespene_used_active_forces, vespene_used_in_progress_economy, minerals_used_in_progress_technology, vespene_used_in_progress_army, food_made, food_used, vespene_friendly_fire_army, vespene_killed_technology, minerals_used_in_progress_economy, minerals_friendly_fire_technology, vespene_lost_army, workers_active_count, vespene_lost_economy, minerals_lost_economy, minerals_friendly_fire_economy, minerals_killed_technology, vespene_lost_technology, minerals_used_current_army, minerals_killed_economy, vespene_collection_rate, vespene_used_current_army, minerals_used_active_forces, vespene_friendly_fire_technology, vespene_killed_army, minerals_friendly_fire_army, vespene_killed_economy

Acknowledgements We would like to thank the StarCraft II community for sharing their experiences, playing together and discussing key aspects of the gameplay. We extend our thanks especially to Mateusz “Gerald” Budziak, Igor “Indy” Kaczmarek, Adam “Ziomek” Skorynko, Jakub “Trifax” Kosior, Michał “PAPI” Królikowski, Konrad “Angry” Pieszak and Damian “Damrud” Rudnicki.

Author Contributions Conceptualization: Andrzej Białecki; Data curation: Andrzej Białecki, Piotr Białecki; Investigation: Andrzej Białecki; Formal Analysis: Andrzej Białecki, Ashwin Phatak; Methodology: Ashwin Phatak; Software: Piotr Białecki; Supervision: Jan Gajewski, Daniel Memmert; Writing - original draft: Andrzej Białecki; Writing - review and editing: Andrzej Białecki, Jan Gajewski, Ashwin Phatak, Daniel Memmert

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability The data used for analyses is available as supplementary material.

Code Availability The code used for analyses is available as supplementary material.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X. (2015) TensorFlow: large-scale machine learning on heterogeneous systems, <https://www.tensorflow.org/>. Software available from tensorflow.org.
2. Alligulac (2013) Starcraft 2 progaming statistics and predictions. <http://aligulac.com/>. Accessed 05 May 2021
3. Asensio J, Peralta J, Arrabales R, Bedia M, Cortez P, López A (2014) Artificial intelligence approaches for the generation and assessment of believable human-like behaviour in virtual characters. *Expert Syst Appl*:11, <https://doi.org/10.1016/j.eswa.2014.05.004>
4. Barr P, Noble J, Biddle R (2006) Video game values: human–computer interaction and games. *Interact Comput* 19(2):180–195. <https://doi.org/10.1016/j.intcom.2006.08.008>
5. Bednárek D, Krulis M, Yaghob J, Zavoral F (2017) Data preprocessing of esports game records - counter-strike: global offensive:269–276. <https://doi.org/10.5220/0006475002690276>
6. Blizzard (2013) S2 protocol. <https://github.com/Blizzard/s2protocol>. Accessed 05 May 2021
7. Braun P, Cuzzocrea A, Keding TD, Leung CK, Padzor AG, Sayson D (2017) Game data mining: clustering and visualization of online game data in cyber-physical worlds. *Procedia Comput Sci* 112:2259–2268. <https://doi.org/10.1016/j.procs.2017.08.141>
8. Chikish Y, Carreras M, Garci J (2019) eSports: a new era for the sports industry and a new impulse for the research in sports (and) economics?:477–508, ISBN 978-84-17609-23-8
9. Consalvo M, Ess C (2011) The handbook of internet studies. ISBN 9781405185882. <https://doi.org/10.1002/9781444314861>
10. Drachen A, Canossa A (2009) Analyzing spatial user behavior in computer games using geographic information systems. *MindTrek 2009 - 13th international academic mindtrek conference: everyday life in the ubiquitous era*:182–189. <https://doi.org/10.1145/1621841.1621875>
11. Edge N (2013) Evolution of the gaming experience: live video streaming and the emergence of a new web community. *The Elon Journal of Undergraduate Research in Communications* 4(2):33–39
12. Edwards TF (2013) Esports: a brief history. <http://adanai.com/esports/>. Accessed 05 May 2021
13. Erickson G, Buro M (2014) Global state evaluation in starcraft
14. Esport Earnings (2012) Esports earnings: prize money / results / history / statistics. https://www.esportearnings.com/history/2017/top_players. Accessed 05 May 2021
15. Freihaut P, Göritz AS (2021) Using the computer mouse for stress measurement – an empirical investigation and critical review. *Int J Hum Comput Stud* 145:102520. <https://doi.org/https://doi.org/10.1016/j.ijhcs.2020.102520>. <http://www.sciencedirect.com/science/article/pii/S1071581920301221>
16. Gaming B (2018) The history and evolution of esports. <https://medium.com/@BountieGaming/the-history-and-evolution-of-esports-8ab6c1cf3257>. Accessed 05 May 2021
17. Good O (2012) Today is the 40th anniversary of the world's first known video gaming tournament. <https://kotaku.com/5953371/today-is-the-40th-anniversary-of-the-worlds-first-known-video-gaming-tournament>. Accessed 05 May 2021
18. H2O.ai. h2o: python interface for H2O. <https://github.com/h2oai/h2o-3>. Python package version 3.30.0.6. (2020)
19. Hallmann K, Giel T (2018) eSports – Competitive sports or recreational activity? *Sport Manage Rev* 21(1):14–20. <https://doi.org/10.1016/j.smr.2017.07.011>
20. Hutchins B (2008) Signs of meta-change in second modernity: the growth of e-sport and the world cyber games. *New Media Soc* 10(6):851–869. <https://doi.org/10.1177/1461444808096248>
21. Intel (2020) Intel brings esports to pyeongchang ahead of the olympic winter games. <https://newsroom.intel.com/news/intel-brings-esports-pyeongchang-ahead-olympic-winter-games/>. Accessed 05 May 2021
22. Jaderberg M, Czarnecki WM, Dunning I, Marris L, Lever G, Castañeda AG, Beattie C, Rabinowitz NC, Morcos AS, Ruderman A, Sonnerat N, Green T, Deason L, Leibo JZ, Silver D, Hassabis D, Kavukcuoglu K, Graepel T (2019) Human-level performance in 3d multiplayer games with population-based reinforcement learning. *Science* 364(6443):859–865. <https://doi.org/10.1126/science.aau6249>. <https://science.sciencemag.org/content/364/6443/859>
23. Justesen N, Bontrager P, Togelius J, Risi S (2020) Deep learning for video game playing. *IEEE Trans Games* 12(1):1–20. <https://doi.org/10.1109/TG.2019.2896986>
24. Kane D, Spradley B (2017) Recognizing esports as a sport. *Sport J* 19:05

25. Kaytoue M, Silva A, Cerf L, Meira W, Raïssi C (2012) Watch me playing, i am a professional: a first study on video game live streaming. WWW'12 - proceedings of the 21st annual conference on world wide web companion, (June 2009):1181–1188. <https://doi.org/10.1145/2187980.2188259>
26. Kim M-J, Kim K-J, Kim S, Dey A (2016) Evaluation of starcraft artificial intelligence competition bots by experienced human players:1915–1921, 05. <https://doi.org/10.1145/2851581.2892305>
27. Kow YM, Young T (2013) Media technologies and learning in the starcraft esports community. Proceedings of the 2013 conference on computer supported cooperative work - CSCW '13:387, ISSN 2156-2202. <https://doi.org/10.1145/2441776.2441821>. <http://dl.acm.org/citation.cfm?doi=2441776.2441821>
28. Lee D, Kim M-J, Ahn CW (2021) Predicting combat outcomes and optimizing armies in starcraft ii by deep learning. Expert Syst Appl 185:115592. <https://doi.org/https://doi.org/10.1016/j.eswa.2021.115592>
29. Liquipedia (2010) Battle.net leagues. https://liquipedia.net/starcraft2/Battle.net_Leagues#Matchmaking-Rating. Accessed 05 May 2021
30. Liquipedia (2011) Stalemate detection. https://liquipedia.net/starcraft2/Stalemate_Detection. Accessed 05 May 2021
31. Liquipedia (2012) The international. https://liquipedia.net/dota2/The_International. Accessed 29 Jan 2022
32. Liquipedia (2015) Units (legacy of the void). [https://liquipedia.net/starcraft2/Units_\(Legacy_of_the_Void\)](https://liquipedia.net/starcraft2/Units_(Legacy_of_the_Void)). Accessed 05 May 2021
33. Low-Kam C, Raïssi C, Kaytoue M, Pei J (2013) Mining statistically significant sequential patterns. Proc IEEE Int Conf Data Min, (October):488–497, ISSN 15504786. <https://doi.org/10.1109/ICDM.2013.124>
34. Menard S (2011) Standards for standardized logistic regression coefficients. Soc Forces 89(4):1409–1428. <https://doi.org/10.1093/sf/89.4.1409>
35. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S (2019) Pytorch: an imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E, Garnett R (eds) Advances in neural information processing systems 32:8024–8035, Curran associates, Inc. <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
36. Ramos G, Meek C, Simard P, Suh J, Ghorashi S (2020) Interactive machine teaching: a human-centered approach to building machine-learned models. Hum Comput Interact 35(5-6):413–451. <https://doi.org/10.1080/07370024.2020.1734931>
37. Ranked For Teh Win (2013) Starcraft 2 ladder rankings. <https://www.rankedit.com/>. Accessed 05 May 2021
38. Reitman JG, Anderson-Coto MJ, Wu M, Lee JS, Steinkuehler C (2020) Esports research: a literature review. Games Cult 15(1):32–50. <https://doi.org/10.1177/1555412019840892>
39. Sánchez-Ruiz AA, Miranda M (2017) A machine learning approach to predict the winner in starcraft based on influence maps. Entertain Comput 19:29–41. <https://doi.org/10.1016/j.entcom.2016.11.005>
40. Sarlis V, Tjortjis C (2020) Sports analytics – evaluation of basketball players and team performance. Inf Syst 93:101562. <https://doi.org/10.1016/j.is.2020.101562>
41. Schubert M, Drachen A, Mahlmann T (2016) Esports analytics through encounter detection. MIT Sloan Sports Analytics Conference:1–18
42. Si C, Pisan Y, Tan CT, Shen S (2017) An initial understanding of how game users explore virtual environments. Entertain Comput 19:13–27. <https://doi.org/10.1016/j.entcom.2016.11.003>
43. Sozański H, Sadowski J, Czerwiński J (2013) Podstawy teorii i technologii treningu sportowego Tom 1, vol 1, AWF Warszawa, Wydział Wychowania Fizycznego i Sportu w Białej Podlaskiej
44. Taylor TL (2012) Raising the stakes : e-sports and the professionalization of computer gaming. MIT Press, Cambridge, Mass. ISBN 9780262017374
45. Thiel A, John JM (2018) Is sport a 'real' sport? reflections on the spread of virtual competitions. Eur J Sport Soc 15(4):311–315. <https://doi.org/10.1080/16138171.2018.1559019>
46. Thompson JJ, Blair MR, Chen L, Henrey AJ (2013) Video game telemetry as a critical tool in the study of complex skill learning. PLoS ONE, 8(9), ISSN 19326203. <https://doi.org/10.1371/journal.pone.0075129>
47. Uriarte A, Ontañón S (2018) Combat models for rts games. IEEE Trans Games 10(1):29–41. <https://doi.org/10.1109/TGIAIG.2017.2669895>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Andrzej Białecki¹  · **Jan Gajewski**²  · **Piotr Białecki**³ · **Ashwin Phatak**⁴  ·
Daniel Memmert⁴ 

Andrzej Białecki
andrzej.bialecki94@gmail.com

Daniel Memmert
D.Memmert@dshs-koeln.de

¹ Warsaw University of Technology, pl. Politechniki 1, 00-661, Warsaw, Poland

² Józef Piłsudski University of Physical Education in Warsaw, Marymoncka 34, 00-968, Warsaw, Poland

³ Bellevue, WA, USA

⁴ Institute of Exercise Training and Sport Informatics, Am Sportpark Müngersdorf 6, 50933, Cologne, Germany