Workshop Proceedings of the 19th International Conference on Intelligent Environments (IE2023)
G. Bekaroo et al. (Eds.)

doi:10.3233/AISE230013

An Exploratory Data Analysis for League of Legends Professional Match Data

Juan MORALES-GARCÍA [a,1], Antonio LLANES [a], Manuel CURADO [a] and Francisco ARCAS-TÚNEZ [a]

[a] *Universidad Católica de Murcia (UCAM), Murcia, Spain*

ORCiD ID: Juan Morales-García https://orcid.org/0000-0003-0008-4825, Antonio Llanes https://orcid.org/0000-0002-9802-4240, Manuel Curado https://orcid.org/0000-0003-2307-1760, Francisco Arcas-Túnez https://orcid.org/0000-0003-4892-5902

Abstract. League of Legends (LoL) is a multiplayer online battle arena video game developed and published by Riot Games. It is a team-based game with over 140 characters to make epic plays with. The game blends the speed and intensity of an real-time strategy game (RTS) with role-playing game (RPG) elements. Two teams of powerful champions, each with unique designs and play styles, battle head-to-head across multiple maps and game modes. Exploratory data analysis (EDA) is a statistical technique that can be used to analyze this data to extract valuable information for both researchers and players. By using EDA techniques on LoL match data, players can identify patterns, trends, and relationships that can help optimize their gameplay strategy. EDA can also help players identify their strengths and weaknesses and important statistics for their gameplay.

The paper provides an introduction to the treatment of LoL match data using EDA techniques. It presents the most common data analysis techniques and explores some examples of how to apply these techniques to LoL match data. Furthermore, the paper discusses some ways in which data analysis can help LoL players improve their game, such as identifying their strengths and weaknesses, patterns and trends, important statistics, and meta changes.

Keywords. League of Legends, LoL, Professional data, Exploratory Data Analysis, EDA

1. Introduction

League of Legends (LoL) is a multiplayer online battle arena (MOBA) video game genre developed by Riot Games with a large community of players worldwide and a international professional league. In LoL, players are formed into two even teams primarily comprised of five members. Each team starts at opposing sides of a map, near what is called a 'Nexus'. To win a match, a team must destroy the opposing team's Nexus. To do so, each team must work through a series of towers called 'turrets' that are placed along three paths to each base (commonly referred to as 'lanes') Along the way, each player gains power by completing game objectives, earning them experience points and gold

[1]Corresponding Author: Juan Morales-García, email: jmorales8@ucam.edu.

that are used to increase the player's level and to purchase powerful items, potentially giving players an advantage over their opponents. Examples of these objectives include killing the opposing team's turrets, players, and 'minions' (small NPCs that constantly spawn and attack the other team).

Like many other MOBA, LoL matches generate large amounts of data in each game that can be analyzed to extract valuable information and knowledge specially valuable for professional team coaches. Exploratory data analysis (EDA) is a statistical technique used to explore and analyze data in order to identify relevant patterns, trends, and relationships. The use of EDA techniques on LoL match data can provide a deeper understanding of the game and help players optimize their gameplay strategy.

This paper provides an introductory study to the treatment of LoL match data using EDA techniques. The most common data analysis techniques will be presented and some examples of how to apply these techniques to LoL match data will be explored. In addition, some ways in which data analysis can help professional LoL teams improve their gameplay will be proposed. In this respect, how can data analysis help LoL players?

- *Identification of strengths and weaknesses*: Data analysis can help LoL players identify their strengths and weaknesses in the game. By analyzing data from previous matches, players can identify which champions and strategies have worked best for them in the past and adapt their gameplay accordingly.
- *Identification of patterns and trends*: Data analysis can help LoL players identify patterns and trends in the game. For example, analyzing data from previous matches can help players identify the most popular champions and roles at a given time, which can be useful for adjusting their gameplay strategy. They can also identify the most effective strategies for defeating certain champions, which can be useful for developing their own gameplay strategy.
- *Identification of important statistics*: Data analysis can also help LoL players identify the most important statistics for their gameplay. For example, players can analyze data to determine which statistics are most important for their particular champion or role, such as attack speed, cooldown reduction, or critical strike chance. By focusing on these important statistics, players can optimize their gameplay strategy.
- *Identification of meta changes*: Data analysis can help LoL players stay up-to-date on changes in the game's meta. The meta refers to the strategies, champions, and items that are most effective at a given time in the game. By analyzing data from previous matches, players can identify trends in the meta and adjust their gameplay strategy to stay one step ahead.

In conclusion, analyzing LoL match data can provide valuable information for players looking to improve their gameplay. By using EDA techniques, players can identify patterns, trends, and improvement opportunities in their gameplay, as well as stay up-to-date on changes in the meta. As a result, they can adjust their gameplay strategy and make more informed decisions on the battlefield.

The rest of the paper is structured as follows. Section 2 shows related work that focuses on similar research work of this paper. Section 3 shows the tools used for the development of this proposal. Section 4 shows the results obtained and the discussion of these results. Finally, Section 5 presents the main conclusions and discusses future works.

2. Related work

League of Legends is a very popular professional video game around the world [1]. This game has famous leagues and tournaments with important awards, making it one of the most relevant video games in the e-sports [2]. Recent research is looking at different ways of making any kind of gain, as the prediction of the winning team in professional matches of the game from previous game data [3,4]. Other studies, such as [5], propose to improve the accuracy of the models trained with the features offered by the game API building a data collection with the skills of a player, the team-synergies or the ability of a gamer. However, its method is far from optimal results (0.70 of classification accuracy).

A different approach [6] investigates how the way of communication between players based on different characters is associated with players' champion and how they are controlled, using information from the game developer and the subjective champion characteristics. The results of champions role are associated with gender, vocality, toxicity, and negative valence.

Using data from the players of the League of Legends team of UCAM e-sports [7], it is analyzed the differences between the pre and post-game mood of the participants to establish variables that affect the behaviour of the players, such as anxiety or self-confidence.

According to the current research state of this research line and the associated problems, we propose a starting point for the treatment of LoL match data using EDA techniques, exploring the most common data analysis techniques and showing some examples of how to apply these techniques to LoL match data.

3. Materials and Methods

This section will show the materials and methods used in this research; specifically, (1) the professional data sources from which the datasets were extracted and (2) a description of the extracted dataset.

3.1. Professional match data sources

Bayes Esports[2] is a company that offers official professional live data, directly from the source of matches and training results through the use of a payment API that allows downloading of JSON files with all the information of a game, using endpoints, such as:

- POST /login and /login/refresh_token: Returns a bearer access token valid during 24h.
- GET /riot-lol/teams: Gets the list of all LoL teams that participated in tournaments that the current authenticated user has access to.
- GET /riot-lol/leagues: Returns a list of leagues and its tournaments for the LoL game.
- GET /riot-lol/tournaments: Returns information of tournaments info and a list of related matches.
- GET /riot-lol/matches: Returns information of each match and its games.

[2] https://docs.bayesesports.com/docs-historic/api

- GET /riot-lol/games: Returns the information for each game in a massive downloadable JSON file or in separate files.

Community Dragon[3] is an open source organisation that provides tools to extract data from Riot's content distribution network, as well as hosting services for extracted data, which is stored in a GitHub repository from which basic information about items, champions and other basic aspects of the game that change with each patch.

Leaguepedia[4] is an API used to get basic information on official matches, which allows to retrieve information about players, teams, tournaments, and matches.

Data Dragon[5] is the official API that allows, among other things, to obtain information about champions and items a few days after each new update of the game. This repository allows to update automatically all this kind of data in a simple way.

Finally, Riot's official API[6] provides free non-professional match information with limitations on the number of requests per time and the level of detail of the data obtained. Due to these limitations and the real motivations of amateur players, more related to leisure than to competition, Bayes Esports, the first of the above data sources, is a better source of game results.

For any artificial intelligence application of a professional LoL gaming environment, a large amount of data collected from real games in professional competitions held on professional LoL servers is needed, and this data is provided by Bayes Esports. Supplementing this data and keeping up to date with the ever-changing hyperparameters of the game engine is possible thanks to Community Dragon, Leaguepedia and Data Dragon sources. With these data sources we can tackle AI projects using Machine Learning and Deep Learning algorithms.

3.2. Data description

A typical game data dump in JSON format is more than 200 MB in size and contains more than 700,000 events. Each event has different fields depending on the type, but transforming the original dump into a table format, 339 unique columns of data appear. Some data fields are of no interest because they keep the same value, fields like /version, always 2, /title, always "LOL", or /payload/type, always "LOL_RIOT_WEB_LIVESTATS", are therefore completely useless to apply in any IA process. Similarly, other fields describe a particular game, indicating the game ID, league name and ID or the ID of the two teams involved.

However, other fields are related to the time sequence of the game, fields like /seqIdx, /timeSent, /createdAt or /startTime are interesting to be used for time series analysis. Finally, the most interesting variables are categorical, such as /action with values as ANNOUNCE, BANNED_HERO, SELECTED_HERO, UPDATE_POSITION, LEVEL_UP, KILL, DIED, STAR_MAP, END_MAP, among other 20 discrete values, but also global numerical variables such as /assists, /baronKills, /championsKills, /dragonKills, /inhibKills, /towerKills, /totalGold or /deaths, as well as other variables related to each character's position, level, armour or skill power.

[3] https://github.com/communitydragon/docs/blob/master/assets.md
[4] https://lol.fandom.com/wiki/Help:Leaguepedia_API
[5] https://developer.riotgames.com/docs/lol#data_dragon
[6] https://developer.riotgames.com/apis

4. Evaluation and Discussion

In this section we will show the data collected from the professional APIs and an exploratory study of the most relevant data for the analysis of the game; specifically, this exploratory study will focus on two aspects: (1) the selection of champions and (2) the events of the game.

4.1. Champions selection

In this subsection, a set of graphical representations will be shown in order to help the understanding of the champion selection process in a professional LoL game.

Figure 1 shows statistics of the data referring to the selection of champions. Specifically, Figure 1a shows a QQ plot to observe the distribution of champion selection events; Figure 1b shows a histogram of the subjects performing the actions, i.e., team or individual players; Figure 1c shows all available actions, i.e., game updates, champion selection, champion ban, and game announcements. Both histograms show the distribution of the two most important discrete variables for understanding the type of event recorded in the JSON file and from which the other most significant variables for the type of event and the origin of the event are derived.

Once the champion selection process has been completed, the teams are established. As an example, Figure 2 shows the final state of a formed team.

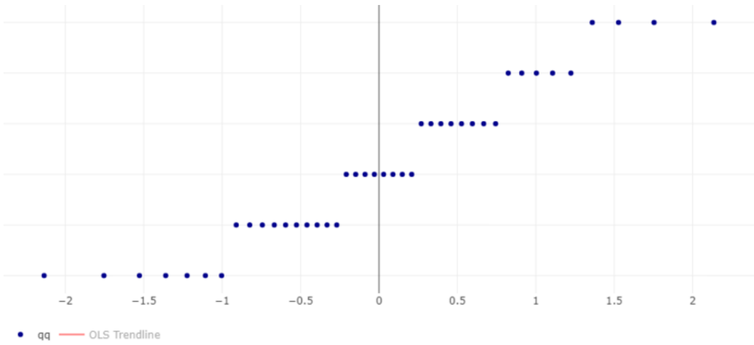
With all the data on the choice shifts, subjects of the action, available actions and final result, a detailed study of the evolution of the selection of champions can be made to evaluate or detect patterns in some of the processes studied.

4.2. Game events

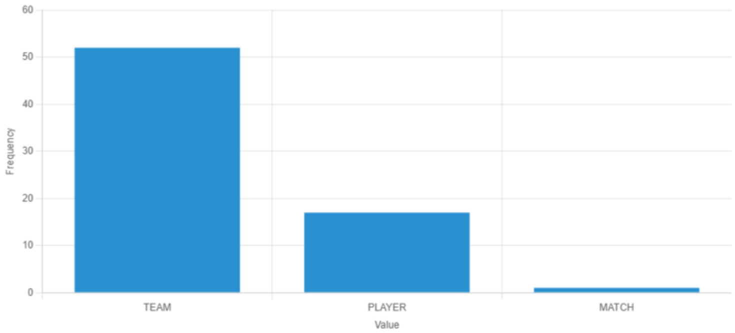
In this subsection, a series of graphical representations will be shown in order to help the understanding of the course and events in a professional LoL game.

Figure 3 shows the most important game events represented by a box-and-whisker plot; in particular, assists (see Figure 3a), deads (see Figure 3b), champions kills (see Figure 3c), dragon kills (see Figure 3d), Baron kills (see Figure 3e) and inhibitor kills (see Figure 3f) are shown.

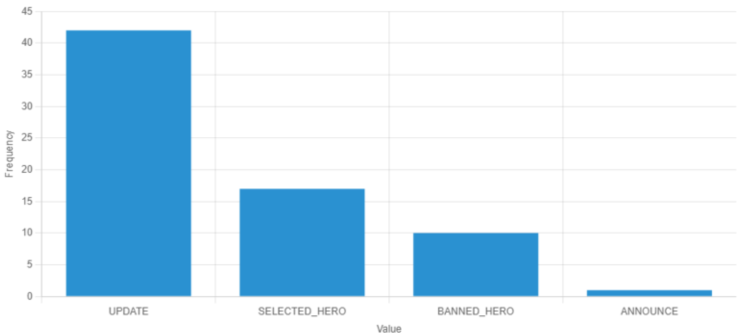
Figure 4 shows the correlation between all the variables previously studied. As can be seen, there is a high correlation between most of them, in fact, the dynamics of the game is observable. For example, Assist is related to ChampionKills (0.97) although a Champion can be killed by an opponent without assistance; The Baron is a passive NPC until a player from either team attacks him, at which point he repels the attack, if he is defeated, all living members of the defeating player's team get Gold, an attack buff for 3 minutes, and his minions also get a temporary boost to all their stats, which is why it is related to the rest of the Kill metrics (0.84, 0.74) as it is one of the best times to go on the attack, especially against enemy headquarters (0.91, 0.95); The relationship between Assist (0.56), ChampionKills (0.44) and TotalGold (0.72) with Deaths, has to do with during team fights since both teams have deaths, also in the winning team, which means Gold for both teams because a Death is a Kill of the other team; Towers protect Inhib, destroying a tower makes it easier to destroy its associated Inhib (0.91); When a champion is successfully killed, it is free to attack the tower directly as there is no one



(a) QQPlot of champions pick turn



(b) Histogram of subjects of the action



(c) Histogram of available actions

Figure 1. Statistical data of champions selection

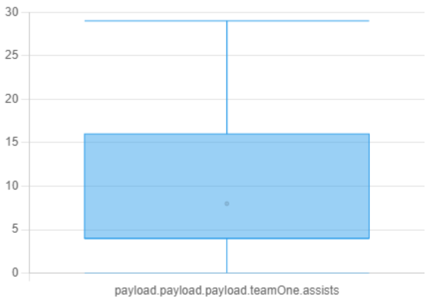
to defend it (0.89). These are just some of the intricate relationships that occur during a match, AI tools such as Graph Neural Networks can be used effectively to model these complex interactions.

Figure 5 shows a dendogram relative to all the variables contained in the dataset under study. This dendogram shows the relationships between all the variables as well as the distances (similarities) between them by means of a graphical representation in the form of hierarchical clustering.

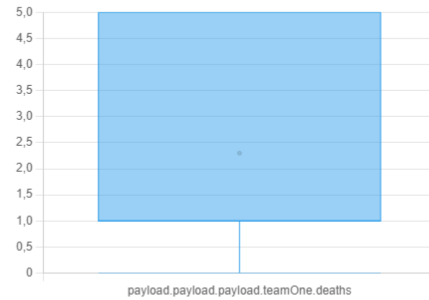
Figure 6 shows the movements of the champions of the two teams within a game.

[{'liveDataPlayerUrn': 'live:lol:riot:player:2614304f-09c4-3d5e-878f-80d4eea5f010', 'liveDataTeamUrn': 'live:lol:riot:team:98767991853197861', 'references': {'RIOT_ESPORTS_ID': '105320680474347057'}, 'championID': 110, 'summonerName': 'T1 Zeus', 'pickTurn': 0, 'pickMode': 2}, {'liveDataPlayerUrn': 'live:lol:riot:player:a660808f-6504-3562-a0aa-38fdfec6d5e6', 'liveDataTeamUrn': 'live:lol:riot:team:98767991853197861', 'references': {'RIOT_ESPORTS_ID': '105320682452092471'}, 'championID': 777, 'summonerName': 'T1 Oner', 'pickTurn': 0, 'pickMode': 2}, {'liveDataPlayerUrn': 'live:lol:riot:player:9ec59e62-aa63-3ec6-a029-ee53fe6780f9', 'liveDataTeamUrn': 'live:lol:riot:team:98767991853197861', 'references': {'RIOT_ESPORTS_ID': '98767991747728851'}, 'championID': 888, 'summonerName': 'T1 Faker', 'pickTurn': 0, 'pickMode': 2}, {'liveDataPlayerUrn': 'live:lol:riot:player:77986919-c81b-331c-a824-743d85805703', 'liveDataTeamUrn': 'live:lol:riot:team:98767991853197861', 'references': {'RIOT_ESPORTS_ID': '103495716775975785'}, 'championID': 0, 'summonerName': 'T1 Gumayusi', 'pickTurn': 2, 'pickMode': 1}, {'liveDataPlayerUrn': 'live:lol:riot:player:b0d56891-47ec-3b1c-964b-2a7915a5fa98', 'liveDataTeamUrn': 'live:lol:riot:team:98767991853197861', 'references': {'RIOT_ESPORTS_ID': '103495716561790834'}, 'championID': 0, 'summonerName': 'T1 Keria', 'pickTurn': 3, 'pickMode': 0}]

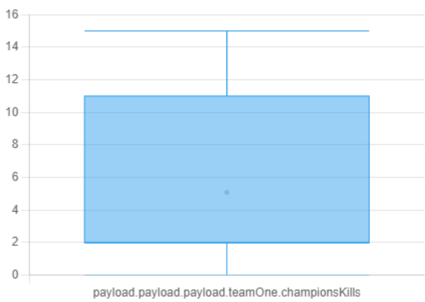
Figure 2. Data of selected champions



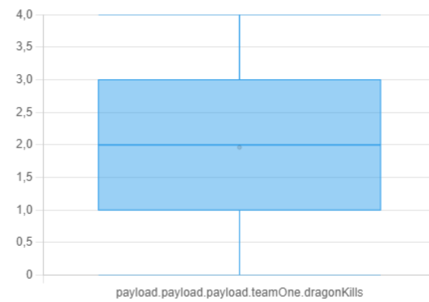
(a) BoxPlot of assists



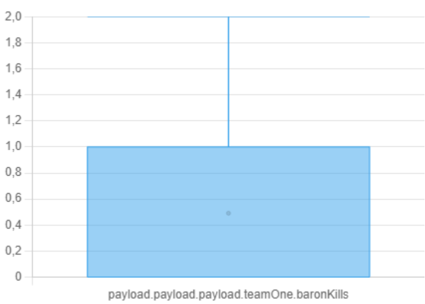
(b) BoxPlot of deads



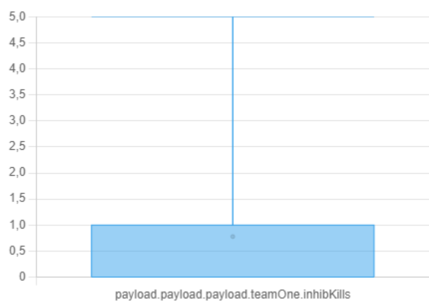
(c) BoxPlot of champion kills



(d) BoxPlot of dragon kills



(e) BoxPlot of Baron kills



(f) BoxPlot of inhibitor kills

Figure 3. Boxplot of the most important game events

It can be seen how the champions of each of the teams advance; in particular, the first

	Assists	BaronKills	ChampionsKills	Deaths	DragonKills	InhibKills	TotalGold	TowerKills
Assists	1,00	0,74	0,97	0,56	0,82	0,64	0,91	0,78
BaronKills	0,74	1,00	0,84	0,11	0,74	0,91	0,64	0,95
ChampionsKills	0,97	0,84	1,00	0,44	0,84	0,73	0,87	0,89
Deaths	0,56	0,11	0,44	1,00	0,11	0,12	0,72	0,11
DragonKills	0,82	0,74	0,84	0,11	1,00	0,67	0,70	0,80
InhibKills	0,64	0,91	0,73	0,12	0,67	1,00	0,59	0,91
TotalGold	0,91	0,64	0,87	0,72	0,70	0,59	1,00	0,67
TowerKills	0,78	0,95	0,89	0,11	0,80	0,91	0,67	1,00

Figure 4. Correlation between studied variables

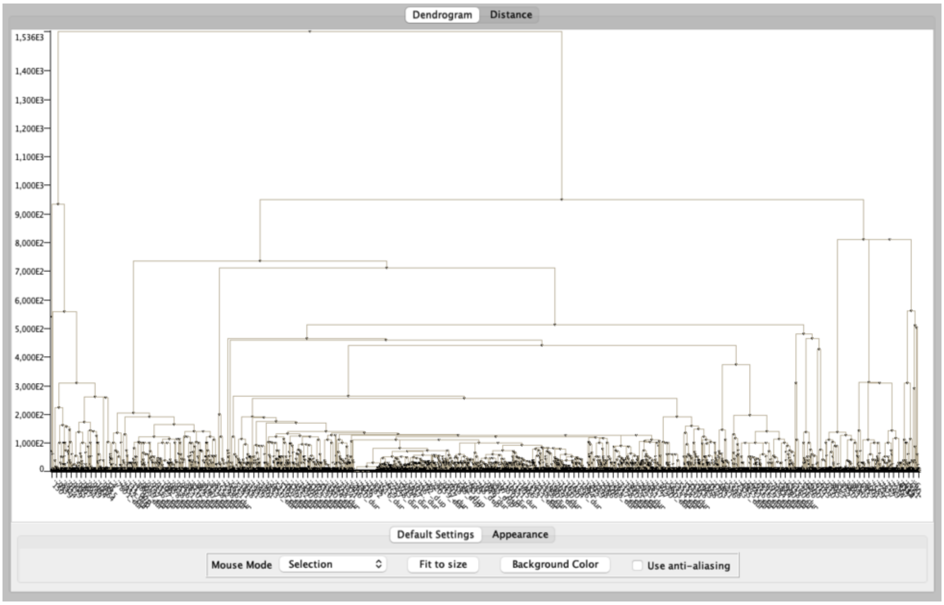
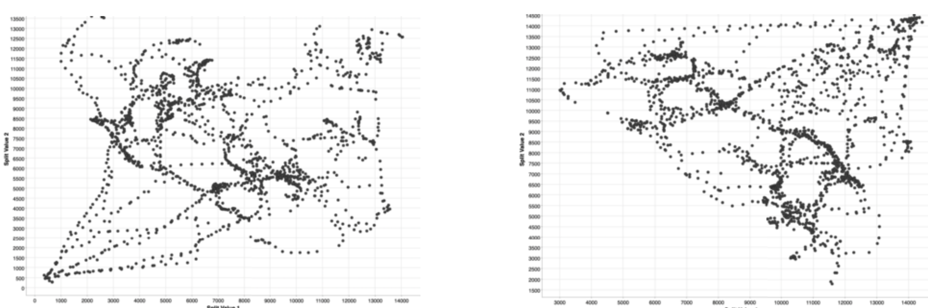


Figure 5. Dendogram of all existing variables in the studied dataset

team (see Figure 6a) advances to the nexus of the rival team, defeating it. On the other hand, the second team (see Figure 6b) does not advance beyond half of the playing field, limiting itself to defending its towers and nexus, although it finally loses the game.



(a) Team one movements

(b) Team two movements

Figure 6. Movements of each of the champions of the two game teams

Thanks to all these aforementioned statistics, it is possible to observe the evolution of a game and all the events that take place within it, to evaluate the performance of each of the teams and, subsequently, to make predictions about who could win the game, to carry out strategies in real time depending on how the scenario in question evolves, among other possibilities.

5. Conclusions and Future work

As we have seen, this paper is a preparatory work to be able to enter the world of the application of Artificial Intelligence techniques in order to predict, explain and improve the best strategies. It is a necessary step to understand the variables, to perform an exploratory analysis of the data so that, in future work, we can focus on more advanced techniques of Artificial Intelligence applied to the data generated by LoL games.

Therefore, the main future work of this proposal lies in the implementation of Artificial Intelligence techniques to make predictions on the dataset to: (1) help the manager's decision making prior to an important match; (2) predict the most powerful champions of the game; (3) search and match champions with in-game objects in order to maximize their base statistics; (4) help the choice of the champion based on the intrinsic characteristics of the player; and (5) build strategies based on the team you are going to face.

Acknowledgments

This work is derived from R&D project RTC2019-007159-5, funded by MCIN/AEI/ 10.13039/501100011033, " FSE invest in your future" and "ERDF A way of making Europe". Financial support for this research also has been provided under grant PID2020-112827GB-I00 funded by MCIN/AEI/10.13039/501100011033.

References

[1] Kicurovski F. Is League of Legends dying in 2023?; 2023. Accessed: 2023-3-14. https://www.esports.net/news/is-league-of-legends-dying/.

[2] Block S, Haack F. eSports: a new industry. In: SHS Web of Conferences. vol. 92. EDP Sciences; 2021. p. 04002.

[3] Silva ALC, Pappa GL, Chaimowicz L. Continuous outcome prediction of league of legends competitive matches using recurrent neural networks. SBC-Proceedings of SBCGames. 2018:2179-259.

[4] Kim DH, Lee C, Chung KS. A confidence-calibrated moba game winner predictor. In: 2020 IEEE Conference on Games (CoG). IEEE; 2020. p. 622-5.

[5] Hitar-Garcia JA, Moran-Fernandez L, Bolon-Canedo V. Machine learning methods for predicting league of legends game outcome. IEEE Transactions on Games. 2022.

[6] Şengün S, Santos JM, Salminen J, Jung Sg, Jansen BJ. Do players communicate differently depending on the champion played? Exploring the Proteus effect in League of Legends. Technological Forecasting and Social Change. 2022;177:121556.

[7] Mateo-Orcajada A, Abenza-Cano L, Vaquero-Cristóbal R. Analyzing the changes in the psychological profile of professional League of Legends players during competition. Computers in Human Behavior. 2022;126:107030.