Recommendation System for Defining Competitive Teams in eSports: Experimental Analysis of CBLOL

Vinícius Antonio Ramos Zecca[1], **Rodrigo Colnago Contreras**[1,2]

[1]University of São Paulo (USP)
São Carlos - SP, 13566-590, Brazil

[2]Federal University of São Paulo (UNIFESP)
São José dos Campos, SP, 04023-062, Brazil

contreras@unifesp.br

Abstract. *Roster building in esports is a complex strategic challenge that can be optimized using data-driven methodologies. This study proposes a novel approach for assembling League of Legends teams in the CBLOL, integrating performance analysis and machine learning techniques. The method accounts for role-specific performance metrics and introduces a network-based metric inspired by the Erdős Number to assess a player's competitive experience. The core objective is to build a recommendation system that supports roster selection by identifying players likely to contribute to team success. Case studies show the methodology's application in team selection, When predicting team performance in the CBLOL 2024 Split 2, the model yielded a Pearson correlation of 0.89 between predicted and actual rankings, highlighting its effectiveness in competitive roster building.*
Keywords *esports; recommendation; machine learning; League of Legends; feature selection.*

1. Introduction

Esports have become one of the most dynamic and lucrative entertainment industries globally, attracting millions of viewers and generating substantial revenue annually. In Brazil, the *"Campeonato Brasileiro de LoL"* (CBLOL) stands out, reaching peaks of over 450,000 concurrent viewers during the April 2024 finals [Mais Esports 2024].

LoL (LoL) is a Multiplayer Online Battle Arena (MOBA) game where two teams of five players compete to destroy the opponent's main structure, known as the "Nexus" [League of Legends 2024]. The game's complexity lies not only in the technical skills of its characters but also in strategic map control, positioning, teamwork, and adaptation to dynamic match conditions. This combination makes team formation a challenge, where selecting the most technically skilled players does not necessarily guarantee success [TechTudo 2020], a common fact in most Esports according to [Nagorsky e Wiemeyer 2020].

Throughout its more than ten-year competitive history in Brazil, CBLOL has undergone significant evolution in team-building strategies. Initially, professional teams prioritized recruiting players based on technical skills, known as "mechanical players" [Kim et al. 2024]. However, teams composed solely of such players did not always achieve the expected success. Several so-called "super teams," which recruited former professionals or high-performance foreign players, often failed to deliver outstanding results [ge.globo 2023a].

In contrast, teams that emphasized a balance between technical ability and behavioral aspects have been more successful. For example, LOUD, a team that joined the league in 2021, dominated the competitive scene from mid-2022 to April 2024. Unlike traditional super teams, LOUD focused on player synergy and cohesion, exceeding expectations from sports analysts and media at the time [ge.globo 2023b].

This phenomenon raises questions about the factors that truly influence a team's performance in professional tournaments like CBLOL. What makes a player more likely to succeed? What characteristics should be prioritized when building a competitive roster? The answer here is not limited to the player statistics by itself, but also the combination of players and different roles in the team.

This study leverages data from "Leaguepedia", a community-driven database that compiles detailed statistics on professional LoL players, teams, and matches, to identify factors influencing roster success. The goal is to develop a recommendation model that suggests players for a given roster by integrating game statistics with a qualitative metric inspired by the Erdős Number [Hoffman 1998]. This metric quantifies a player's relevant experience based on competitive history, offering a way of measuring seniority.

The model is adapted to each in-game position, ensuring that success metrics align with the specific expectations of each role. The approach aims at providing coaches and team managers with a tool for optimizing roster composition for competitive Esports tournaments, using CBLOL as an example of executing the framework.

2. Material and Methods

2.1. Database

The dataset was extracted through the Leaguepedia API, which allows direct access to its "Cargo Tables" via structured queries [Fandom 2024]. These tables contain historical and updated data on professional leagues, including CBLOL, Worlds and other tournaments.

The dataset was built by retrieving records from the *ScoreboardPlayers* and *ScoreboardTeams* tables. The first table provides match-level player statistics such as champion selection, kills, deaths, assists, vision score, and gold earned, while the second includes team compositions and results. To establish relationships between players, data from both tables were merged, associating each competitor with their teammates and opponents across multiple matches. Additionally, aggregated statistics per match were extracted from the *ScoreboardGames* table to incorporate broader team-level metrics.

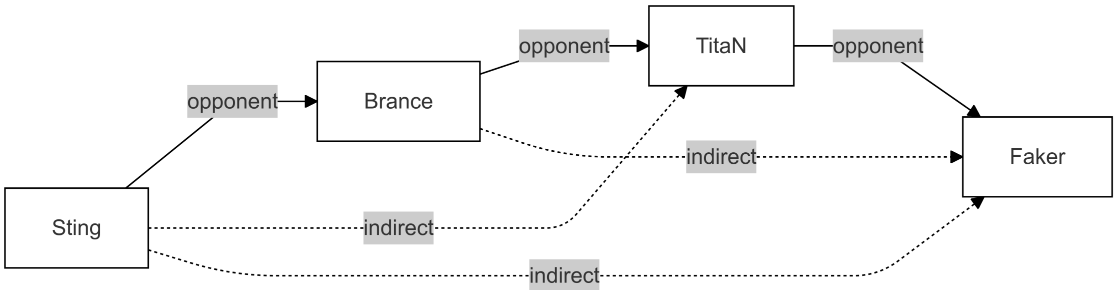
To ensure consistency, preprocessing steps were applied, including data cleaning, duplicate removal, and structuring the dataset for analysis. The final dataset consists of two components: a player relationship network, mapping connections based on competitive interactions, and a statistical dataset containing individual and team performance metrics.

2.2. Player Relationship Graph

To model interactions between players, an undirected graph $G = (V, E)$ was constructed, where each node $v \in V$ represents a player, and each edge $e \in E$ denotes a competitive interaction. Edges connect players who have either been teammates or opponents in official matches, forming a network that captures relational proximity in the competitive matches space. This approach aligns with the framework presented by [Kearns et al. 2001], where multiplayer games are modeled using undirected graphs to represent local interactions among players.

The shortest path distances between all player pairs were computed using the Breadth-First Search (BFS) algorithm without any weights applied, meaning the distance equals the number of edges between one player and another [Moore 1959]. This approach identifies both direct and indirect connections, enabling the quantification of player proximity.

Figure 1. Example of the Network, connecting "Sting" and "Faker".



The resulting graph provides a structured representation of player relationships, serving as a foundation for feature extraction. In this work, a new feature was created representing the distance between each player and Lee "Faker" Sang-hyeok, considered the best player of all time in LoL, as the first one added to the game's Hall of Fame [Esports 2024]. This measure helps quantify seniority of CBLOL players, since, in theory, the smaller the distance to Faker, the higher is their international and relevant competitive experience. Figure 1 illustrates an example of this structure.

2.3. Composite Score Calculation

To measure a player's performance, a composite score was developed, integrating individual statistics with dynamically assigned weights. Instead of applying fixed weights to each metric, the methodology determines their relevance based on historical matches, ensuring that each metric's contribution aligns with the role of the player.

Weights were determined using logistic regression models trained separately for each in-game role. By modeling the probability of winning the game based on performance statistics, this approach identifies which metrics have the greatest influence on match outcomes for different positions. The relative importance of each feature is determined by the regression coefficients (β_i), which represent the strength of the association between each metric and the probability of victory [Hosmer e Lemeshow 2000, Xu et al. 2016]. These coefficients have been widely used in predictive modeling to assess feature significance in different domains, including player performance and competitive environments [La Cava et al. 2019].

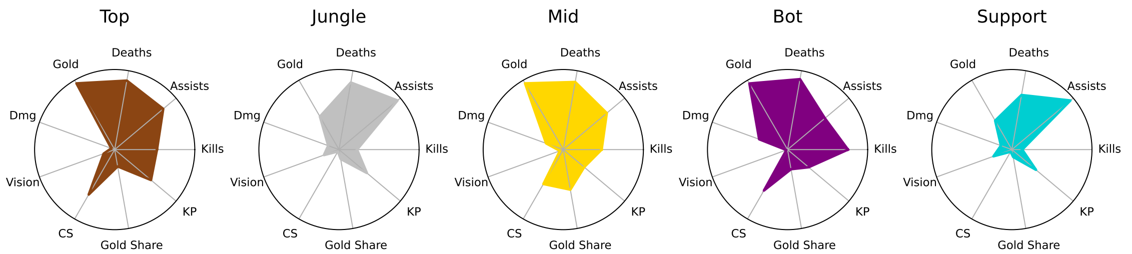
To ensure comparability and reduce the influence of metric scale differences, all feature values were standardized prior to weight application. This standardization minimizes bias caused by varying ranges and ensures that the weights derived from the logistic regression reflect relative importance rather than raw magnitue of the metric since they follow different scales. The resulting coefficients are then normalized to form a probability distribution, where the sum of the weights equals 1. This allows direct use of the weights in the calculation of the composite score, accurately reflecting the relevance of each metric in both individual and team performance.

The final composite score of a player is computed as:

$$\text{Composite Score} = \sum_{i=1}^{n} w_i \cdot x_i, \tag{1}$$

where w_i represents the weight assigned to the i-th metric, and x_i is the corresponding metric value in its original scale [OECD 2008]. This process ensures that the derived score effectively summarizes a player's overall impact based on relevant performance indicators.

Figure 2. Normalized distribution of feature importance per in-game position. Each bar represents the relative importance of a given metric for predicting match victory, with all weights normalized to sum to 1 for each role.



Source: Author (2025).

For instance, "Vision Score" is more significant for Supports, while "Kills" and "Damage to Champions" play a larger role for Bot laners. This can be visualized in Figure 2, where the importance of each metric is represented as a normalized weight derived from the logistic regression models trained per role. The weights are scaled to form a probability distribution, ensuring comparability across roles. This reinforces the necessity of using role-specific weighting in player evaluation, as it ensures that the composite score accurately reflects each player's contribution relative to their position in the game.

Before training, all metrics were standardized to ensure comparability, and the dataset was split into training and testing subsets for evaluation. Segmenting the model by role allowed it to capture role-specific performance patterns, improving the accuracy of weight assignments while reflecting what matters to their position in the game.

2.4. Recommendation Model Design

The recommendation model was trained on aggregated player data, incorporating individual performance metrics, the composite score, and the "Faker Number" as predictors, with the Erdős-inspired metric serving as a proxy for relevant competitive experience. Its objective is to estimate a player's likelihood of contributing to a winning team using a multi-class classification approach, distinguishing between players who reached the playoffs (class 1), those who won the championship (class 2), and those who did not qualify (class 0).

A *Random Forest* classifier was used due to its capability to process both categorical and numerical data while capturing complex interactions between variables [Breiman 2001]. The recommendation system operates iteratively, selecting players based on a user-defined role priority, allowing flexibility in roster construction.

The recommendation process begins by incorporating any preselected players and computing a weighted probability of winning and qualifying for playoffs, referred to as the "Winning Potential Score" (WPS). WPS is derived by aggregating the predicted probabilities of team success across a proposed roster, based on the outputs of the classification model. This score represents the expected competitive performance of a candidate lineup. The model evaluates WPS for different team compositions and selects the player that maximizes the score, and this process repeats iteratively for each remaining role until all positions in the roster are filled.

3. Results and Discussion

3.1. CBLOL Split 2 2024

The model's predictive accuracy was tested using data from CBLOL 2024 Split 2, comparing its ranking predictions to real tournament results. The evaluation, summarized in Table 1, demonstrates the model's effectiveness in aligning its rankings with actual team results. The computed WPS positioned most teams within one or two ranks of their final standing, with a Spearman correlation of 0.89 indicating a strong monotonic relationship between predictions and outcomes [Zar 2005].

Table 1. Model predictions for CBLOL 2024 Split 2 compared to actual standings.

Team	Winning Potential Score	Predicted Position	Real Position
paiN Gaming	0.655	1	1
RED Canids	0.645	2	3
FURIA	0.610	3	5
Vivo Keyd Stars	0.550	4	2
LOUD	0.515	5	4
INTZ	0.185	6	8
KaBuM! Esports	0.160	7	7
Fluxo	0.055	8	6
Liberty	0.040	9	9
Los Grandes	0.035	10	10

The model correctly predicted the final standings for paiN Gaming, Liberty, and Los Grandes. Deviations for the remaining teams were small and mostly within one or two ranks. These differences are expected given the presence of external factors such as coaching strategy, roster synergy, or game meta adaptation, which are not captured by individual player statistics. In cases like RED Canids and Vivo Keyd Stars, the ranking inversion reflects marginal score differences rather than model error. Overall, the results show that the model can differentiate team strength with high precision using only pre-match data.

4. Final Considerations

The work proved to be an effective approach for optimizing team formation in competitive Esports, demonstrating that the methodology can enhance roster selection by integrating performance metrics, role-specific weighting, and relevant competitive experience. The main purpose of this work is to propose a new approach that integrates performance statistics, competitive experience, and role-specific modeling into a unified recommendation framework, with an emphasis on actionability rather than descriptive analytics.

Beyond roster management, the methodology can be applied in broader Esports analytics contexts, including tournament forecasting and market evaluations for player acquisitions. It can also be applied for Esports bets and fantasy games, as it uses statistical data from players that can be adapted to other games other than LoL.

As a next step, refining the model for cross-regional applicability could improve how it interacts with other tournaments, by incorporating data from multiple leagues to mitigate "Meta" variations and player skill level. Additional features, such as tracking player online game performance, along with the professional experience, and integrating external factors like social media sentiment, could improve the accuracy of recommendations, specially with the integration of a large language model agent to help interpret and classify qualitative data. These ideas will make the system more adaptable for other competitive environments and other factors out of the game.

5. References

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

Esports, L. (2024). Lol esports welcomes faker to hall of legends. Accessed: 2025-01-05.

Fandom (2024). Leaguepedia - api documentation.

ge.globo (2023a). Cblol já teve supertimes antes da kabum, e eles fracassaram.

ge.globo (2023b). Final cblol 2023: Loud é melhor time da história do brasil, dizem robo e tin.

Hoffman, P. (1998). *The Man Who Loved Only Numbers: The Story of Paul Erdős and the Search for Mathematical Truth*. Hyperion, New York.

Hosmer, D. W. e Lemeshow, S. (2000). Applied logistic regression.

Kearns, M., Littman, M. L., e Singh, S. (2001). Graphical models for game theory. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI2001)*, pages 253–260, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Kim, J., Dunkel, W., e Lee, B. (2024). Operationalizing general mechanical skill in time-pressure action esports. *Journal of Electronic Gaming and Esports*, 2(1):11.

La Cava, W., Bauer, C., Moore, J. H., e Pendergrass, S. A. (2019). Interpretation of machine learning predictions for patient outcomes in electronic health records. *arXiv preprint arXiv:1903.12074*.

League of Legends (2024). Como jogar.

Mais Esports (2024). Cblol 2024: Liga teve o terceiro maior pico de audiência de finais.

Moore, E. F. (1959). The shortest path through a maze. *Proceedings of the International Symposium on the Theory of Switching*, pages 285–292.

Nagorsky, E. e Wiemeyer, J. (2020). The structure of performance and training in esports. *PLOS ONE*, 15(8):e0237584.

OECD (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*.

TechTudo (2020). Cblol 2021: Entenda seleção dos times e regras do sistema de franquias esports.

Xu, E. L., Qian, X., Liu, T., e Cui, S. (2016). Detection of cooperative interactions in logistic regression models. *arXiv preprint arXiv:1602.03963*.

Zar, J. H. (2005). Spearman rank correlation. *Encyclopedia of Biostatistics*.