

# Datasheet for League of Legend Player Performance KDA analysis Dataset\*

Colin Sihan Yang

2024-11-26

## 1 Datasheet for BLG Player Performance Dataset

### 1.1 Motivation

1. *For what purpose was the dataset created?*

The dataset was created to predict player performance for Bilibili Gaming (BLG) in the League of Legends Pro League (LPL). It aims to assist teams in making data-driven decisions on player strategies, performance optimization, and resource allocation.

2. *Who created the dataset?*

The dataset was created by Riot Game and oracleselixir oracleselixir (2024).

### 1.2 Composition

1. *What do the instances represent?*

Each instance represents an individual game played by a BLG player, including in-game metrics and results.

2. *How many instances are there in total?*

The dataset contains approximately 470 instances, corresponding to all games played by BLG in the 2024 LPL season.

3. *Does the dataset contain all possible instances?*

The dataset is a complete record of BLG games for the selected season. However, it does not include games from other teams or non-professional matches.

---

\*Code and data are available at: [https://github.com/hoodiexxx/lol\\_win\\_rate\\_prediction](https://github.com/hoodiexxx/lol_win_rate_prediction).

4. *What data does each instance consist of?*  
Each instance includes metrics such as vision score, total gold, total CS, damage to champions, KDA, game duration, and match outcomes.
5. *Is there a label or target associated with each instance?*  
Yes, the target variable is the Kill-Death-Assist ratio (KDA).
6. *Is any information missing from individual instances?*  
No significant data is missing. All in-game metrics are complete.
7. *Are relationships between instances made explicit?*  
No explicit relationships between instances are defined. Each instance is independent.
8. *Are there recommended data splits?*  
Yes, the dataset is split into training (80%) and testing (20%) sets.
9. *Are there errors, noise, or redundancies?*  
Minor inconsistencies may exist due to variations in game logs, but these have been minimized during preprocessing.
10. *Is the dataset self-contained?*  
Yes, the dataset is self-contained and does not rely on external resources.
11. *Does the dataset contain confidential data?*  
No, all data is derived from publicly available match records.
12. *Does the dataset contain offensive or sensitive data?*  
No, the dataset is free from offensive or sensitive content.
13. *Does the dataset identify subpopulations?*  
Yes, the dataset categorizes players by role (e.g., Top Lane, Jungle).
14. *Is it possible to identify individuals?*  
No, the dataset only includes player names and their roles without sensitive personal information.
15. *Does the dataset contain sensitive data?*  
No, the dataset does not include sensitive attributes like race, religion, or financial data.
16. *Any other comments?*  
None.

### 1.3 Collection Process

1. *How was the data acquired?*  
The data was extracted from publicly available match records and APIs such as Riot Games' data API.
2. *What mechanisms were used to collect the data?*  
Data was collected using API scripts and verified for completeness and accuracy.
3. *If the dataset is a sample, what was the sampling strategy?*  
The dataset is not a sample; it is a complete record of BLG's 2024 matches.
4. *Over what timeframe was the data collected?*  
Data was collected during the 2024 LPL season.
5. *Were any ethical review processes conducted?*  
Ethical approval was not required as the data is publicly available.
6. *Did you collect the data directly?*  
Yes, data was collected from Riot Games' API and match records.
7. *Were individuals notified about the data collection?*  
Not applicable, as the data pertains to publicly available information.
8. *Did individuals consent to data collection?*  
Not applicable, as the data pertains to publicly available information.
9. *Any other comments?*  
None.

### 1.4 Preprocessing, Cleaning, and Labeling

1. *Was any preprocessing done?*  
Yes, data was cleaned to remove missing values and calculate derived metrics like KDA.
2. *Was raw data saved?*  
Yes, the raw data is archived for future use.
3. *Is the preprocessing software available?*  
Yes, preprocessing scripts are available in the accompanying GitHub repository.
4. *Any other comments?*  
None.

## 1.5 Uses

1. *Has the dataset been used for any tasks already?*  
Yes, it was used to build a Bayesian mixed-effects model for predicting player KDA.
2. *What other tasks could the dataset be used for?*  
The dataset could be used for performance analysis, team strategy optimization, and meta-analysis of in-game metrics.
3. *Are there tasks for which the dataset should not be used?*  
The dataset should not be used for any non-esports purposes, such as identifying personal information.
4. *Any other comments?*  
None.

## 1.6 Maintenance

1. *Who will maintain the dataset?*  
No one
2. *How can the dataset owner be contacted?*  
N/A
3. *Will the dataset be updated?*  
N/A
4. *Any other comments?*  
None.

## References

oracleselixir. 2024. “Oracle’s Elixir - LoL Esports Stats.” <https://oracleselixir.com/tools/downloads>.