Ecamania Esports Project - League of Legends

FKY

2025

Figure 1: Ecamania logo

Introduction

This year, a new student-led project has been approved by the administration: the creation of an Esports team. The project received strong support from the director, Mr. Dekimpe, who promised to provide an **UNLIMITED budget**! The team behind this project decided to launch their first season focusing on *League of Legends*, one of the most competitive games. He wants to hire the best players in the world, no matter what the cost, to come play and study at ECAM.

As the data analyst for the newly formed team **Ecamania**, you have been assigned several missions:

1. Determine, based on in-game statistics of a single player, whether their team won or lost the match.
2. Identify the most important quantifiable variables that players should focus on to increase their chances of winning.
3. Determine the most "talented" players Ecamania could recruit for the next season.

Your manager has found a dataset from professional League of Legends games obtained through the Leaguepedia API and that is available in the files game_events.csv, game_metadata.csv, game_players.csv .

This project was inspired by the paper "PandaSkill - Player Performance and Skill Rating in Esports: Application to League of Legends" from *De Bois et al.* [1], you are more than encouraged to read it to grasp every nuance of this project. What is asked from you is a simpler version of their work that is resumed through the methodology in the next section.

Proposed methodology

To succeed in this challenge, we propose that you follow a given methodology. You are, however, free to create or use another one.

- Visualize and understand your dataset.
- Select the features and pre-process the data.
- Train different models to predict a game winner based on one player's performances. For each model type create a different model for each player role (top, mid, jungle, ADC and support). Find the best hyperparameters and model.
- For each player, use your best model to compute the chance of his/her team winning for all the games in the dataset. These probabilities can represent the player performances, the impact it had on the game. Transform these performances based on the percentile of the role to maintain consistency between roles then aggregate in a smart way these values to obtain a skill rank for every player.
- Use the SHAP (SHapley Additive exPlanations) values of the model to determine how the different variables impacted the game output and which one to focus on during training. An alternative could be univariate analysis and correlation between the variable and the game's result.

Practicalities

The project will be evaluated based on a report, **maximum 10 pages**, and the code produced by the student. The report is to be well-structured with different sections corresponding to the different steps of your methodology. Each section should have a clear description of what was done, graphs representing them and a critical discussion of the work done. **It is recommended for the project to be done by groups of two**, but it can also be done alone. Throughout the entire project, keep a critical approach to what you do and discuss your results, don't just take a descriptive approach (avoid empty GPT-like statements that would penalize your grades). The deadline of the submission on claco is on the **20th of May**. **An evaluation grid is available lower, read and use it**.

League of Legends Overview

Game Layout

League of Legends (LoL) is a competitive Multiplayer Online Battle Arena (MOBA) game. In LoL, two teams of five players each compete on a map called *Summoner's Rift*, with the goal of destroying the enemy Nexus, located inside the enemy base and protected by Towers and Inhibitors.

The two bases are connected by three lanes: the **top lane**, the **mid lane**, and the **bottom lane**. At regular intervals, Minions spawn from both bases and push down the lanes, attacking enemy structures. Players assist the Minions to advance towards the enemy Nexus. Between the lanes is the **jungle**, an area filled with neutral monsters. The strongest of these, such as the *Drake* or *Baron Nashor*, grant significant bonuses to the team that defeats them.

Players and Roles

Each player controls a **Champion**, a character with unique abilities, strengths, and weaknesses. Players also assume one of five roles, which define both their area of influence on the map and their responsibilities within the team:

- **Top**: Plays in the top lane; typically durable and independent; provides map control and frontline presence.
- **Jungle**: Roams between lanes; applies pressure across the map; helps secure objectives.

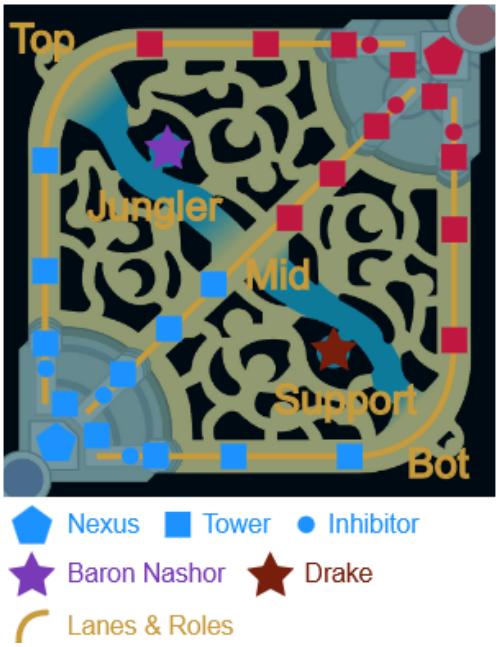

Figure 2: Enter Caption

- **Mid**: Controls the mid lane; centrally located; often the playmaker with high burst damage.
- **Bot (ADC)**: Plays in the bottom lane alongside the Support; provides consistent damage output; scales well into late game.
- **Support**: Also plays in the bottom lane; focuses on vision control and utility; does not require as much gold to be effective.

General Game Strategy

A standard game lasts between 20 and 50 minutes. Players aim to grow stronger by gaining experience and purchasing powerful items with gold earned from killing Minions, monsters, enemy Champions, and destroying structures. Capturing key neutral objectives such as Drakes or the Rift Herald provides significant team advantages. Once strong enough, teams attempt to siege and break into the enemy base.

The Esports Ecosystem

The highest level of competition in League of Legends is organized by the publisher, Riot Games. The world is divided into regions, each with its own league and tournament format. Twice a year, the top teams from each region meet during international tournaments: the *Mid-Season Invitational (MSI)* and the *World Championship (Worlds)*.

Historically, Korea and China have dominated the international scene, followed by Europe and North America. However, as the esports scene matures, emerging regions (e.g., Asia-Pacific) are beginning to pose serious challenges to the established powerhouses.

Evaluation Grid

Legend:

- **Absent** – No evidence of the criterion.

- **Insufficient** – Attempt made but poorly executed or misunderstood.
- **Sufficient** – Meets the basic expectations.
- **Good** – Above average, solid execution.
- **Excellent** – Outstanding, insightful, and well-justified work.

Criteria	**Absent**	**Insufficient**	**Sufficient**	**Good**	**Excellent**
Data Understanding and Analysis (10%)					
Understand dataset structure	Not done	Minimal structure identified	General understanding	Detailed exploration with insights	Deep and thorough understanding
Exploratory Data Analysis (EDA)	Not done	Basic stats only	Basic EDA and visuals	Meaningful insights	In-depth EDA with patterns and trends
Data visualization	Not done	Poor or unclear visuals	Basic graphs	Relevant graphs with analysis	High-quality and insightful visuals
Feature importance analysis	Not done	Poor or unclear analysis	Basic importance discussed	Importance supported with logic or metrics	Advanced techniques well justified
Feature Selection and Creation (15%)					
Relevant feature selection	Not done	Arbitrary or irrelevant features	Some useful features selected	Justified selection using data or domain logic	Optimal features chosen via analysis
Feature transformation / creation	Not done	Irrelevant or redundant features	Basic transformations or new features	Useful engineered features	Innovative and insightful feature creation
Data Pre-processing (10%)					
Missing values / outliers	Not done	Poor or inappropriate handling	Reasonable handling	Good justification	Robust and nuanced approach
Categorical encoding	Not done	Inconsistent or flawed encoding	Basic encoding used	Proper encoding with logic	Advanced or balanced encoding strategies
Scaling numerical features	Not done	Incorrect or partial scaling	Appropriate scaling used	Good choice of method	Method adapted to model needs
Model Comparison and Hyperparameter Selection (25%)					

Dataset splitting	Not done	Incorrect or biased splitting	Basic split (train/test)	Valid train/val/test split	Valid train/val/test split and cross-validation
Model training	No models trained	One or irrelevant model	2 models tested	3 models tested (among which one deep learning model)	more than 3 models tested (among which one deep learning model) with choice justification.
Performance comparison	Not done	Poor metrics or unclear comparison	Basic metrics compared	Structured and meaningful comparison	In-depth multi-metric evaluation
Hyperparameter selection/tuning	Not done	Poor or basic tuning	Basic cross-validation	Good tuning and validation	Advanced tuning with cross-validation
Over/underfitting analysis	Not done	Incorrect or shallow	Basic discussion	Correct analysis with remedies	Deep, justified understanding
Project goals (20%)					
The 3 missions of the project	Not done	1 mission is met	2 missions are met	2 missions are met with correct justification/analysis	The 3 missions are met with correct justification/analysis
Presentation of Results (20%)					
Report structure and visuals	Unreadable	Poorly structured or weak visuals	Basic structure and graphs	Clear structure, good visuals	Excellent clarity and presentation
Methodology and discussion	No method described	Weak or incoherent logic	Acceptable structure and explanation	Coherent and logical methodology	Rigorous and critically discussed methods

Bonus Points (up to 10%) :

- **Innovation:** You explore novel approaches or techniques to improve model performance or address specific challenges of this project.
- **External data source:** Import data from external sources to obtain new meaningful features and merge it in the dataset.
- **Model:** Use unusual models or models that were not seen in the course.

References

[1] Maxime De Bois et al. “PandaSkill-Player Performance and Skill Rating in Esports: Application to League of Legends”. In: *arXiv preprint arXiv:2501.10049* (2025).