UNIVERSIDAD Panamericana

Graphic Simulation

Performance Analysis of teams during League of Legends World Championship 2024

Student: Arturo Garcia Brambila

Professor: Gabriel Castillo Cortés

GitHub Repo

Final Project

Due date: 25/05/2025

Introduction

This project centers on the foundational principles of Machine Learning (ML), specifically applying them to a comprehensive dataset derived from the League of Legends World Championship 2024. Our primary objective was to systematically extract, refine, and clean this data, preparing it for subsequent analytical and predictive modeling tasks. The ultimate aim is to establish a robust analytical framework that can serve as a strong base for future endeavors, potentially allowing for the prediction of tournament outcomes based on intricate team and player performance metrics.

Data Processing

The dataset utilized for this analysis was sourced from Kaggle, offering an extensive collection of statistics from professional matches. This rich compilation includes diverse performance indicators, ranging from gold per minute (GPM) and experience per minute (XPM) for individual players, to details about specific matches and their dates.

The initial phase of our codebase involved the essential importation of necessary libraries, followed immediately by the reading and initial cleaning of the raw data. During this crucial stage, we meticulously processed column names, standardizing them to a more readable format using lowercase letters and underscores for spaces. Furthermore, any columns deemed redundant or irrelevant for the project's scope were systematically removed, streamlining the dataset.

Following the initial cleaning, we proceeded to establish distinct DataFrames to facilitate more organized data management and analysis. A key creation was the match_df DataFrame, which was generated by aggregating all player-level data within a single match into match-specific summaries. Complementing this, a team_df DataFrame was developed to consolidate individual player statistics into team-level performance metrics. This involved summing and averaging player data to produce comprehensive team statistics such as team gold, team experience, gold difference at fifteen minutes (gd@15), team dragon kills (team_dragon), and team baron kills (team_baron). These aggregated DataFrames provide a structured and accessible foundation for deeper analytical dives and subsequent model training.

Data Analysis

With the data meticulously extracted and processed, we established a solid foundation for our analytical exploration. Our initial findings quickly revealed a discernible advantage for the blue team side, evidenced by its consistently higher overall win rate throughout the championship. This statistical leaning suggests inherent strategic or structural benefits associated with starting on the blue side of the map.

Subsequently, our focus shifted to a detailed examination of team performance. We precisely calculated each team's win rate across the entire championship, providing a clear hierarchy of success. Beyond mere victories, we computed the averages of crucial in-game statistics for every team, including kills, deaths, assists, gold earned, and experience gained. This comprehensive statistical profiling offered valuable insights into each team's distinct playstyle, indicating whether they favored aggressive skirmishes, meticulous farming, or objective control.

We also delved into the champion meta, a vital aspect of competitive League of Legends. By analyzing the top 10 most banned champions and the top 10 most picked champions, we identified the characters deemed either overwhelmingly strong and strategically disruptive by opponents, or those considered exceptionally versatile and powerful by teams during the drafting phase. This analysis provided a clear snapshot of the prevailing champion landscape and the strategic priorities of professional players.

Logistic Regression

The goal of this project was to build a machine learning model capable of predicting the outcome of League of Legends matches based on the engineered features. For this binary classification task (win or loss), Logistic Regression was chosen as the primary predictive model.

Logistic Regression is a suitable choice for several reasons: it is a robust and widely understood algorithm, its coefficients offer a degree of interpretability regarding feature importance, and it provides probabilistic outputs for classification. The model fundamentally operates by fitting a sigmoid function to the data, transforming a linear combination of features into a probability score between 0 and 1, which can then be thresholder to predict a class.

This project successfully demonstrated the power of data-driven analysis in the context of professional League of Legends. Several key findings emerged from our exploration:

- The Blue side consistently holds a statistical edge in win probability within this dataset, underscoring a persistent strategic element of the game
- Detailed team performance metrics allowed for the identification of top-tier organizations and the characterization of various team playstyles (e.g., aggressive vs. objective-focused)
- Analysis of champion picks and bans provided a snapshot of the professional meta, revealing which champions were considered dominant power picks or critical strategic threats.
- Novel visualizations like the Radar Chart proved effective in comparing multi-dimensional team performance profiles in a visually intuitive manner.

Future Work and Recommendations

While significant progress has been made, several avenues for future exploration and model enhancement remain:

- Expanded Feature Engineering: The current feature set is robust, but further engineering could enhance model performance. This might involve creating features that capture champion synergies, counter-pick relationships, specific item build timings, or contextual features related to the specific game patch. Incorporating time-series analysis to understand the evolution of gold leads, kill disparities, or objective control throughout a match could also provide rich, dynamic features.
- Deeper Player-Level Analysis: While aggregated team data is powerful, a more granular analysis focusing on individual player performance consistency, champion mastery, or unique player playstyles could uncover additional predictive signals.

Conclusion

Like previous undertakings, this project greatly benefited from the guidance of an AI assistant. This support was instrumental in navigating the Python codebase, effectively complementing my existing skills and facilitating the execution of various analytical and modeling tasks.

This project proved to be a highly engaging and enjoyable experience, particularly because it involved data that I not only understood but had also observed in real-time during the 2024 LoL World Championship. Directly witnessing the tournament unfold provided invaluable context, allowing me to intuitively discern which data points would be most impactful for the project's objectives and which information might be redundant or irrelevant. This practical understanding significantly enhanced my feature engineering process. Furthermore, the hands-on experience of preparing the data and training the model deepened my comprehension of how machine learning fundamentally operates and learns from data to build predictive models.